

AT

DEPARTMENT OF HEALTH AND HUMAN SERVICES

PUBLIC HEALTH SERVICE

FOOD AND DRUG ADMINISTRATION

CARDIOVASCULAR AND RENAL DRUGS ADVISORY COMMITTEE

82nd MEETING

Thursday, October 23, 1997

8:40 a.m.

National Institutes of Health
Clinical Center-Building 10
Jack Masur Auditorium

MILLER REPORTING COMPANY, INC.
507 C Street, N.E.
Washington, D.C. 20002
(202) 546-6666

9000 Rockville Pike
Bethesda, Maryland

PARTICIPANTS

Milton Packer, M.D., Chairperson
Joan C. Standaert, Executive Secretary

MEMBERS

Robert Califf, M.D.
Thomas Graboys, M.D., (Consumer Representative)
John DiMarco, M.D.
Marvin Konstam, M.D.
JoAnn Lindenfeld, M.D.
Lemuel Moye, M.D., Ph.D.
Ileana Pina, M.D.
Dan Roden, M.D.C.M.
Udho Thadani, M.D., FRCP

TEMPORARY VOTING MEMBER

Ralph D'Agostino, M.D., Ph.D. (Chair, OTC
Committee)

FDA

Raymond Lipicky, M.D.
Robert Temple, M.D.

C O N T E N T S

Introductory Remarks: Milton Packer, M.D.	4
Open Public Hearing	
BASIC STATISTICAL CONSIDERATIONS FOR THE EVALUATION OF ACTIVE CONTROLLED CLINICAL TRIALS	
View of Cardiorenal Division Regarding Positive Controlled Trials:	
Robert R. Fenichel, M.D., Ph.D.	7
"If That is Your View," Then This is What you Have to Think About:	
Rory Collins, M.D.	55
"If These are the Circumstances," This is How to Calculate Things:	
David DeMets, Ph.D.	94
General Discussion	127

1 P R O C E E D I N G S

2 DR. PACKER: We will be beginning the 82nd Meeting
3 of the Cardiorenal Drugs Advisory Committee. This
4 particular meeting marks the first appearance of some new
5 members on the Committee. So, in order to introduce the new
6 members and also introduce some of the invited guests for
7 today's meeting -- Mark, do you want to start? Just
8 introduce yourself and institution of origin.

9 DR. KONSTAM: Marv Konstam, New England Medical
10 Center, Boston.

11 DR. LINDENFELD: JoAnn Lindenfeld, University of
12 Colorado.

13 DR. RODEN: Dan Roden, Vanderbilt University.

14 DR. PACKER: Milton Packer, Columbia University.

15 DR. PINA: Ileana Pina, Temple, Philadelphia.

16 DR. CALIFF: Rob Califf, Duke University.

17 DR. MOYE: Lem Moye, University of Texas Health
18 Science Center, Houston.

19 DR. GRABOYS: Tom Graboys, Brigham and Women's
20 Hospital, Harvard Medical School.

21 DR. THADANI: Udho Thadani, University of
22 Oklahoma.

23 DR. D'AGOSTINO: Ralph D'Agostino, Boston
24 University.

1 DR. PACKER: Dr. D'Agostino is a temporary voting
2 member for today and tomorrow's meetings. We also have the
3 courtesy and the privilege of having two guest experts at
4 today's meeting, Dr. Rory Collins from the University of
5 Oxford and Dr. David DeMets from the University of
6 Wisconsin. We will be hearing from both of those experts in
7 a short time.

8 Joan, do you want to read the conflict of interest
9 waivers and other administrative issues for today's and
10 tomorrow's meeting?

11 MS. STANDAERT: I will do the one for today; I
12 will do another one tomorrow. The following announcement
13 addresses the issue of conflict of interest with regard to
14 this meeting, and it is made a part of the record to
15 preclude even the appearance of such at this meeting.

16 The purpose of this meeting is to have a general
17 scientific discussion of basic statistical considerations
18 for the evaluation of active controlled clinical trials.
19 Since no questions will be addressed to the Committee by the
20 Agency on issues dealing with a specific product, IND, NDA
21 or firm, it has been determined that all interests and firms
22 regulated by the Center for Drug Evaluation and Research,
23 which have been reported by the participants present,
24 present no potential for a conflict of interest at this

1 meeting when evaluated against the agenda. However, in the
2 event that the discussions involve any products or firms not
3 on the agenda for which an FDA participant has a financial
4 interest, the participants are aware of the need to exclude
5 themselves from such involvement and their exclusion will be
6 noted for the record.

7 With respect to all other participants, we ask in
8 the interest of fairness that they address any current or
9 previous financial involvement in any firm whose products
10 they may wish to comment upon.

11 That concludes the conflict of interest statement
12 for October 23, 1997.

13 DR. PACKER: Thank you very much. It is
14 traditional to reserve time at the beginning of each day for
15 public comment. Is there any public comment? There not
16 being any public comment today, we will proceed with the
17 primary objective in the agenda for today's meeting.

18 The purpose today is to have a broad-based
19 discussion on statistical considerations in the evaluation
20 of active controlled clinical trials. This is intended to
21 be a broad-based overview and exploration of issues. There
22 are no questions that will be posed to the panel, and there
23 may or may not be any conclusions reached by the panel. The
24 idea is to identify issues and try to explore them as best

1 as one can.

2 Let me simply advise the Committee that today's
3 deliberations really should be separated from tomorrow's
4 deliberations so that we should make every effort today not
5 to specifically refer to any issues about tomorrow's meeting
6 in today's discussion.

7 With that in mind, the first presentation will be
8 by Dr. Robert Fenichel, who will present the view of the
9 Cardiorenal Division regarding positive controlled trials.

10 Ray, are you supposed to give an introduction?

11 DR. LIPICKY: No.

12 **Basic Statistical Considerations for the Evaluation**
13 **of Active Controlled Clinical Trials**
14 **View of the Cardiorenal Division**

15 **Regarding Active Controlled Clinical Trials**

16 DR. FENICHEL: Dr. Packer, members of the
17 Committee, ladies and gentlemen, good morning.

18 We are going to talk about active controlled
19 trials this morning and about the possibility of using them
20 to draw conclusions about how a procedure would have
21 performed had it been present.

22 (Slide)

23 As you will see on this slide, we have been

1 talking about these trials as the putative-placebo trials as
2 an alternative to various other nomenclatures used, and I
3 will talk about why the nomenclature is what it is. The
4 mode of analysis is something that we started talking about
5 in the Division in December of 1992 when this Committee was
6 discussing thrombolytic agents and, as mentioned here, at
7 least two other times at meetings of the Committee but never
8 really fully explained and it keeps changing and developing.
9 So some of what I am going to say repeats what was said in
10 '92; some can be said to be similar to some recent documents
11 from the ICH and other sources; but some is quite new and
12 may be idiosyncratic to the Division.

13 (Slide)

14 Let me start where we usually start, classic
15 superiority trial. We have to contrast the putative-placebo
16 trial to this sort of thing. This is the most familiar sort
17 of trial. The object is to show that the test drug is
18 different from, but one hopes superior to some control,
19 usually placebo but not necessarily.

20 If you do a sloppy trial, then you may not see the
21 difference and sloppiness can be as simple as statistical
22 sloppiness where the sample size is not large enough to tuck
23 in the confidence interval, but it could be something else
24 like not really knowing how to take blood pressures, not

1 making sure the patients get drug they are assigned to,
2 other things that will not show up in your statistical
3 analysis.

4 The control in a superiority trial may be an
5 active drug. It doesn't have to be placebo but it is harder
6 to win against an active drug. The new valuable drug may
7 not be more effective than any particular active drug. It
8 may be safer or cheaper, or better tasting, whatever. And
9 even if it is more effective, it may be only so slightly
10 more effective that that is very hard to show without a
11 prohibitively large trial. So, there are some difficulties
12 with classical superiority trials but they seem like the
13 straightforward way to demonstrate superiority to placebo
14 which is, after all, the criterion of provability in the
15 United States.

16 (Slide)

17 So, why do we look at any other design? Well,
18 this is a street scene in Pokhara, Nepal. Let me focus in
19 on this. Sometimes you can't do placebo-controlled trials.

20 (Slide)

21 So you have to do something else.

22 (Laughter)

23 (Slide)

24 Well, there is the notion of a classical

1 equivalence trial. The classical equivalence trial is
2 successful if the outcome of the test drug is
3 indistinguishable from the outcome of the active control.
4 Well, the problem here -- and Dr. Temple and others have
5 been writing about this for ten or fifteen years -- is that
6 the easiest way to be indistinguishable from the active
7 control is to be indistinguishable from anything -- just to
8 be so sloppy that you are not taking blood pressure
9 actively; that you don't really care who got which drug and
10 so, of course, everything comes out the same. Noise is on
11 your side. And some of this will be apparent if your sample
12 size is simply so small that you couldn't tell the
13 difference, but some of it will not be apparent in poor
14 execution and design of the trial.

15 So, FDA has been fairly hostile to this sort of
16 trial and we have moved on in our thinking to the idea of
17 the putative-placebo trial.

18 (Slide)

19 First of all, when do we consider these? Well, we
20 say consider them when placebo would be an unethical sort of
21 trial to conduct as a superiority trial. And that is a
22 community criterion which is not necessarily even agreed
23 upon by FDA in a given case, but there it is. Some trials
24 cannot be done. Then other situations are that one is

1 simply not likely to win against that kind of control.

2 Remember, the new drug may not be as good but may still be
3 good enough. It may be, as I say, cheaper, better tasting,
4 or whatever.

5 Finally, the most important fact about this is
6 that there is a known active control which can be used,
7 which is so consistently superior to placebo that
8 performance with respect to it can be a gauge of performance
9 against the placebo which is not there.

10 That is not always the case, that there is such an
11 active control. In some clinical areas, even where we
12 believe that the existing drugs are effective and have
13 proved those existing drugs, the existing drugs may
14 frequently unpredictably fail to manifest their efficacy.
15 Analgesics do that; antidepressants do that. But in other
16 situations there will be active controls with reliable
17 magnitudes of effect.

18 The other thing, which I have stuck on the very
19 last line of this slide, is that there is a known active
20 control with these desirable properties is an FDA judgment,
21 as contrasted to the possibility that placebos would be
22 impossible, which is a community judgment, and sometimes
23 there may be a community judgment that placebos can't be
24 used but an FDA judgment that active controls, appropriate

1 active controls are not known. That is a difficulty. It
2 has implications way beyond this agenda and that is another
3 talk.

4 DR. LIPICKY: Bob, can we interrupt you?

5 DR. FENICHEL: Yes.

6 DR. LIPICKY: I am not sure I understand the
7 differentiation you are making between FDA judgment and
8 community judgment. What do you mean by those words?

9 DR. FENICHEL: To give an example, there is
10 community judgment which we have been resisting with mixed
11 success, I think -- well, pretty good success, that it is
12 unethical to do placebo-controlled trials of antianginals.
13 Now, it is also true that I don't think there is -- I think
14 that is just plain wrong, as many people in the audience
15 will know and certainly the members of the panel will know.
16 A retrospective review of antianginal trials shows that as
17 regards safety it is a little bit safer to be on placebo in
18 antianginal trials, as submitted to the Agency, but that is
19 not known to many IRBs and so it might be true that it
20 simply can't be done, to do a placebo-controlled trial of an
21 antianginal. Nevertheless, we would say that there is no
22 active control which performs so well -- we might say this,
23 that performance against it of such-and-such a magnitude,
24 such-and-such an efficacy would be convincing as evidence of

1 efficacy against placebo. That would put the sponsor, of
2 course, in a great bind, but that is the bind I am referring
3 to here.

4 DR. LIPICKY: But do you mean that something has
5 to be an approved drug to be an active control?

6 DR. FENICHEL: No.

7 DR. LIPICKY: So, since FDA has not made a
8 judgment and approved it, that is not the judgment you are
9 referring to here?

10 DR. FENICHEL: No, that is not the judgment. It
11 would be a different kind of judgment. But it would be
12 difficult in a non-approved case to say exactly how we would
13 come to that judgment. As I said, that really is another
14 talk and the time is somewhat short.

15 (Slide)

16 Putative-placebo trials -- the idea is that the
17 trial is successful if the outcome with the test drug is
18 superior, not necessarily to the active control but to the
19 best outcome that might have been seen with placebo if
20 placebo had been present. The point about this is this is a
21 superiority trial; it is not an equivalence trial. So, if
22 you do a sloppy trial, if your sample is too small, if you
23 are not really measuring blood pressure or if you give the
24 wrong drugs to patients, then you won't find the

1 superiority. You will lose.

2 The other thing which I have listed here in
3 parentheses is, is the effect size adequate? And we are not
4 really sure what adequate means, and I will come back to
5 that. This came up in the initial discussion on putative-
6 placebo trials when we contemplated the possibility of a new
7 thrombolytic that might be definitely better than placebo
8 but with only a small fraction of the benefit of, say,
9 streptokinase. This effect size adequate clause was meant
10 to capture the idea that if the fraction was small enough
11 then it really might not be appropriate to approve a drug
12 for this purpose. And I will return to this because this is
13 a very big issue here.

14 (Slide)

15 Let me get back to the preconditions. Is a known
16 active control consistently superior to placebo? What does
17 that mean?

18 (Slide)

19 Well, here is a drawing of this. We have a new
20 trial in which the efficacy of the control is up here. Here
21 is the active control. And we can estimate -- this is the
22 quality of the control that we need, which many potential
23 controls do not possess -- we can estimate with confidence
24 that placebo, had it been present, would have been down

1 here, plus/minus this, and the upper bound of placebo
2 confidence limits would have been substantially worse than
3 where this thing, otherwise, plainly, we don't know what we
4 are doing.

5 This confidence limit is a confidence limit that
6 is not about the position of placebo; it is a confidence
7 limit of this estimate of the difference. I will return to
8 that, which is very important. I have another slide which
9 is almost the same thing.

10 (Slide)

11 Here is a shorthand version of the previous slide.
12 Here is this confidence limit but it is not a confidence
13 limit of placebo performance. It is derived from the
14 variance of the placebo active difference. The significance
15 of that is that we need an active control in the new trial
16 because the historical data does not necessarily give us a
17 reproducible value for the efficacy of placebo or of
18 control, for that matter. The historical data that I am
19 looking for give us a reproducible difference between the
20 two efficacies.

21 The example of that which we keep referring to
22 because it was the one that we started thinking about this
23 with was looking at post-infarction use of thrombolytics.
24 Active thrombolytics at that time had been compared to

1 placebo in multiple trials in the course of about a decade
2 and survival in the placebo group had been improving all
3 that time because of concomitant therapy. It ranged from
4 about 88% or so in the old trial up to about 93% in the new
5 trials. So there is no sense in talking about a good
6 estimate for the mortality in placebo groups or, for that
7 matter, in active control groups. Those were both moving
8 targets.

9 It did make sense, when we looked at this five
10 years ago and I don't know if it still make sense, to talk
11 about the difference between the placebo group mortality and
12 the thrombolytic group mortality. That was fairly stable.
13 It ran about 2.5% plus/minus about 0.5%. So, there is the
14 2.5% and there is the 0.5%, more or less.

15 (Slide)

16 What does this mean in practice? Well, here are
17 three examples, and we used the numbers from the
18 thrombolytic example. The numbers are barely visible on the
19 slide and some of them got cut off. But suppose that the
20 active control survival in the new trial is 95%, which is
21 better than it has ever been but that could be because all
22 sorts of concomitant stuff is improving, and so on. So, the
23 argument is, well, the active would have been somewhere
24 around here; the placebo would have been here; and the best

1 would have been here; and a very bad placebo would have been
2 around there.

3 Here is A, which is the test drug that we used,
4 and we know that it is worse than placebo could have been.
5 Here is B. B looks pretty good. It is probably better than
6 the active control but, on the other hand, it might be no
7 better than placebo. So, we would say that it flunks,
8 although I should say that this is shorthand. If this
9 happened again and again you would say, well, suppose this
10 happened ten times and ten times it is numerically better
11 than the putative placebo, that is 1(-10). That is pretty
12 impressive. But in a one trial case we can certainly say
13 that flunks.

14 Here is C. C is worse than the active control.
15 That is what that means, that little gap. But it is better
16 than placebo could have been. We are going to get back to
17 the effect size question which is important, but it passes
18 the first test. C was better than placebo could have been.

19 Well, these are only three examples. There are
20 only twenty possible examples and I am going to show you a
21 rat's next slide.

22 (Slide)

23 There it is. It is a very busy slide. The only
24 point of the slide is to show you that it is possible to go

1 through all of the qualitatively different cases. It takes
2 about five or ten minutes to write them all down and to see
3 them.

4 The other thing which is nice about this slide is
5 that almost all of these cases flunk, with the same standard
6 of flunk that I used before. All of these flunk. Here is A
7 again, worse than placebo could possibly have been. Here is
8 R which really looks pretty good. It is better than the
9 active control but not probably better than placebo could
10 have been. So, they all flunk. There is not much left.

11 (Slide)

12 Here is what is left from two slides ago. T, this
13 thing here, is another easy case. It sounds easy but I want
14 to say just a couple of things about that. First of all, it
15 beats the active control. It must be approvable on efficacy
16 grounds if the control is. But I want to go into T a little
17 bit because it is not the slam-dunk that you think, and I
18 want to say something about the intuition as a guide in this
19 area.

20 (Slide)

21 Here is a mathematical banality: if A is bigger
22 than B and B is much bigger than C, then A must be much
23 bigger than C. Well, that is pretty boring. You don't need
24 a very precise definition of "much bigger" to see that that

1 is true.

2 The second thing just seems like a straight
3 analogy, A beats B in some trial and B crushes C. B equals
4 0.00001. Then A must certainly beat C fantastically. Well,
5 that is not true, or not necessarily true.

6 (Slide)

7 Here is B beating C, four standard deviations,
8 with B on the order of 10^{-4} .

9 (Slide)

10 Here is what people tend to think about in this
11 kind of thing. Here is A beating B, by now only two
12 standard deviations. So, that B is around 0.05. Here is
13 this 0.0001 again. You combine the two; you get six
14 standard deviations. This isn't even tabulated. I had to
15 get this by approximation. So, it sounds like that old
16 inequality that gave you a minute ago. What is wrong with
17 that? Well, what is wrong with that is that is not the only
18 way you can draw these factors together.

19 (Slide)

20 Here is the same story. Here is B beating C, four
21 standard deviations, 0.0001. Here is A beating B, two
22 standard deviations but they are big standard deviations. A
23 is pretty good but you did a small trial. Here is A beating
24 C, 0.04. That is better than 0.05 but it is not that much

1 better. So, this is just a demonstration that intuition may
2 not be your best guide in this area. You have to draw a lot
3 of pictures and go through a lot of stuff.

4 (Slide)

5 I want to get back to the others and say, okay,
6 they turned out to be better than placebo could have been.
7 Is that good enough? So, now we are talking about this.
8 So, now we know what this first diamond is about and we
9 really made no use of the outcome associated with the active
10 control. We have used it as a tool to get into the idea of
11 where placebo is but we really have to look at the effect
12 size.

13 (Slide)

14 It is worth saying that effect size is not often
15 given great weight by the FDA. The Committee and the
16 Division have sometimes grumbled about the marginal size of
17 a demonstrated effect but the Committee and the Division
18 have rarely, if ever, failed to approve a product that beat
19 placebo. Over the six years or so during which we have
20 accumulated placebo-controlled trial data regarding use of
21 ACE inhibitors for congestive failure we never forced
22 sponsors to do comparative trials. Now, we might have said,
23 gee, here's this new drug and, from its trial it looks like
24 the effect is not that terrific. Why don't you do a head-

1 to-head with enalapril? Surely, if it's preserving just a
2 small fraction of enalapril's health-giving, life-giving
3 effect, it should not be approved. Well, we never did that,
4 or in the other direction. I don't mean to be partial to
5 enalapril. But you ought to do a head-to-head. Maybe
6 enalapril shouldn't be approved any more. But we didn't do
7 that.

8 Certainly, antihypertensive packages often include
9 one or two active control trials and sometimes the new drug
10 loses in the active control trials, but we don't especially
11 penalize it. Other regulatory jurisdictions do pay somewhat
12 more attention to comparative results. Maybe we should.
13 But that is a separate question from the question of active
14 control trials.

15 That argument is okay. Nevertheless, when we are
16 using an active control to determine where the new drug
17 falls with respect to placebo, the comparative data are
18 really there right in our faces. So, perhaps we shouldn't
19 discard them, and perhaps it is not a uniform policy but
20 there they are. Also, we are doing these active control
21 trials usually because a placebo would be unethical. What
22 that means -- one way of putting that is it is unethical to
23 expose subjects to zero percent of the effect of established
24 therapy. If 0% is not ethical, is 1% ethical? Is 5%, 50%?

1 So, there is some reason to think about effect size, but
2 before deciding what percentage to use and before picking a
3 number we have to know what is being compared to what, and
4 there are several possible things one can do here.

5 (Slide)

6 One thing you can do, certainly, is simply
7 estimate the drug effect. We know where the test drug came
8 out. We know where placebo would have been, or we have an
9 estimate of where placebo would have been and we can talk
10 about the drug effect, which is here.

11 But the other thing we can do is say how much do
12 we really know we have? After all, that being bigger than
13 zero is what made us decide that this is certainly not a
14 placebo and certainly okay. So, we can look at the
15 guaranteed drug effect.

16 There is a reminder in this drawing that we got
17 this guaranteed effect from the drug effect by subtracting
18 two standard deviations here and another two standard
19 deviations here. So, we are talking about a very small
20 number often. Ordinary superiority trials are thought to
21 have succeeded when something is bigger than two standard
22 deviations and here you have four. So, there are those two
23 measures that we want to talk about, and you can make
24 various comparisons.

1 (Slide)

2 We have another complicated slide, but it is
3 really sort of a duplication of the previous slide in that
4 we have the same thing here -- well, we have an UFO here; I
5 don't know what that is on the slide, but we have the same
6 thing here: the test drug with its guaranteed effect and the
7 test drug effect and we have the control which, in this
8 case, is a little bit worse than the test drug point
9 estimate and it also has a control effect and the guaranteed
10 effect. Now we can say, well, what should be a fraction of
11 what? What are we talking about here?

12 (Slide)

13 This slide puts some numbers on it. It is getting
14 sort of verbose so I just put A, B, C, D going across here
15 so we can refer to these things without going through
16 verbiage. Well, there are at least these four
17 possibilities. You can require that the test drug
18 guarantees at least half as much, or whatever you want to
19 say, of the guaranteed effect of the control, or a quarter
20 or whatever. You can compare the drug effect of one to
21 another. Some of these comparisons seem a little bit more
22 meaningful than others.

23 (Slide)

24 This just repeats what I said a minute ago. Each

1 guaranteed drug effect calculation incorporates four
2 standard deviations. So, these are always small and it is
3 hard to find comparisons of this guaranteed effect versus
4 drug effect that the control would reliably pass against
5 itself, which can lead to some fairly paradoxical sort of
6 results which are not desirable from a regulatory point of
7 view.

8 (Slide)

9 So let me show two examples which I obtained,
10 frankly, by drawing things at random and then measuring them
11 with a ruler. But I think these are practical possible
12 examples. This is the same one I used before and now I have
13 put some numbers in. Here is the guaranteed effect of the
14 new drug, and it is, you know, only about 40% of this
15 guaranteed effect of the old drug. You could also say,
16 well, what we estimate the drug effect does is that it does
17 is a little bit better than the control.

18 You can also make these other comparisons. You
19 can say, well, its guaranteed effect is a very tiny fraction
20 of what we think the control drug effect is. The other
21 thing you can say about the test is, gee, its effect is
22 almost twice what the guaranteed effect of the other one is.
23 These two middle comparisons are not especially meaningful,
24 it seems to me. It is very hard to understand what is going

1 on. This seems to provide some information. This seems to
2 provide a little information. These two things in the
3 middle are problematic.

4 (Slide)

5 The last example is a sort of singularity example.
6 The test drug is exactly the same as the control. It may be
7 a new formulation of the same drug, or something else. But
8 you have a fairly small trial and so the confidence limits
9 are wide. The guaranteed effect of this new thing, you
10 don't know as well what you are getting. That seems like a
11 fair comment. You know that as a point estimate you are
12 getting the same thing exactly. That is a fair comment.
13 This figure and this figure are difficult to interpret, to
14 say the least.

15 (Slide)

16 Where do we go from here? The putative-placebo
17 trials are a new entity. They have been called equivalence
18 trials but that is a misnomer because they succeed when they
19 find a difference, not when they fail to find one. They
20 have been called non-inferiority trials. That is another
21 misnomer because a trial might be successful despite being
22 inferior to the active control, and non-inferiority to
23 placebo would certainly never be adequate. They are
24 impossible without adequate reliable active controls, and

1 those controls will be unavailable in many areas.

2 As we conceive them, the statistical standard for
3 these trials is a difficult one to meet. Possibly, however,
4 this is a one-trial standard often so maybe the one-trial
5 standard at four standard deviations is maybe not a whole
6 lot more demanding than the historical demand of two sigma
7 trials at a conventional level of significance.

8 Finally, the last point here is that we do have
9 some tentative suggestions as to the vocabulary and
10 calculations, the descriptions of the adequacy of effect
11 size, although I am uncertain that this issue of effect size
12 is any more closely connective to active control trials than
13 to others.

14 Thank you for your attention.

15 DR. PACKER: Bob, don't go away. Any comments or
16 questions from the Committee?

17 DR. TEMPLE: The new nomenclature is interesting.
18 I just want to mention some things about the old
19 nomenclature just so it doesn't disappear. After many years
20 of calling these kinds of trials equivalency trials which,
21 as Bob says, is certainly a misnomer, we have somewhat
22 grappled our way to calling them non-inferiority trials.
23 Bob is correct, it is a slight misnomer but I would argue
24 that it is not too much of a misnomer.

1 There are two international guidance documents
2 under development to describe these situations and to
3 address the very problem that Bob has been describing.
4 Failing to show a difference between two treatments is not
5 very informative. There are a lot of reasons for failing to
6 show a difference.

7 In non-inferiority terms what people try to do is
8 identify a difference between drug and placebo that can
9 regularly be shown by a control drug. Bob suggested that a
10 way to do that is to look for the confidence interval. I am
11 not sure that is actually sufficient. Sometimes you can
12 have a confidence interval that describes a difference and,
13 yet, many trials of the drug might not beat placebo at all,
14 even though on average they do. My favorite example for
15 that would be beta-blocker post-infarction trials where I am
16 sure you could draw a confidence interval for the 35 or 40
17 trials that have been carried out and, yet, probably 30 out
18 of those 35 trials didn't distinguish drug and placebo.
19 Whether that reflects the population, sample size or
20 whatever is hard to know. So I would argue that this is
21 something like the case that Bob described before where no
22 one would feel ethically comfortable doing a post-infarction
23 placebo-controlled beta-blocker trial and, yet, an active
24 control trial would be uninformative because you couldn't

1 describe the difference that could regularly be shown
2 between drug and placebo.

3 Anyway, one way people have described these things
4 is that in an active control trial where you are trying to
5 demonstrate not exactly equivalence or non-inferiority but
6 that some effect is there, what you do is you set a margin
7 for the difference between control drug and placebo that, if
8 exceeded or if the 95% confidence interval exceeds it, would
9 tell you, you have not met your standard for non-
10 inferiority. Now, Bob is correct, you can be slightly
11 inferior and, yet, be better than placebo but that is a very
12 unlikely occurrence.

13 So, internationally these are being called non-
14 inferiority trials. We will have to consider what Bob says
15 and maybe abandon that. But the difficulty is setting the
16 margin that represents an amount of difference between drug
17 and placebo that could always be distinguished by the test
18 drug.

19 DR. CALIFF: I have two concerns, and others will
20 keep coming up during the course of the day, but two
21 concerns about the methodology. First, you are in essence
22 talking about historical controls. You are assuming that
23 what is observed in a comparative group somehow is going to
24 be reflective of what was observed in previous direct

1 comparisons when the therapeutic environment is pretty
2 dramatically changing in a lot of the disease states that we
3 treat. So I worry about that.

4 The second worry is the non-inferiority paradigm
5 is almost completely clinically irrelevant. One of the
6 concerns I hope will be discussed in some detail today is
7 what the risk is to the public health of flooding the market
8 with therapies which were shown to be better than putative
9 placebo, but get on the market without any way of the
10 clinician knowing how it compares to the current standard
11 treatment. You know, I would at least take a position right
12 now that we should better try to focus on real equivalence
13 trials and figuring out how to do them right, comparing new
14 treatments with treatments that are known to work, rather
15 than trying to meet some artificial regulatory standard
16 which doesn't really help the patients who are being served.
17 I guess that is a little bit controversial.

18 DR. FENICHEL: Well, I will let either my boss or
19 my boss's boss respond, both of whose hands are up.

20 DR. TEMPLE: Rob, there is no such thing as true
21 equivalence. All you can do is say the difference is not
22 larger than a certain amount. What Bob has described is one
23 way of describing what that difference that you have to
24 exclude should be. If all you want to know is that the drug

1 is better than nothing, then you must exclude a difference
2 that represents the difference between drug and placebo that
3 is guaranteed. If, as Bob said, you say, no, that is not
4 good enough; there is a mortality effect and I need to
5 preserve 75% it, then you set a margin such that if the
6 difference between drug and placebo is potentially greater
7 than 25% of that difference you say, no, I won't approve it.
8 But there is no such thing as equivalence.

9 DR. CALIFF: I am not arguing with the concept
10 that we need to define what the minimally important
11 difference is. I just think we are using the wrong
12 comparative. I think the comparative should be the active
13 therapy which is known to benefit patients --

14 DR. LIPICKY: It is.

15 DR. TEMPLE: It is the active therapy.

16 DR. CALIFF: No, really what you are doing is
17 comparing it to a putative difference with a putative
18 placebo.

19 DR. LIPICKY: No --

20 DR. TEMPLE: No, that is how you are interpreting
21 the active control. You have the active control there and,
22 as a practical matter, if you are not almost as good as the
23 active control numerically you will never exclude the
24 difference you want to exclude.

1 DR. CALIFF: Well, we will keep coming back to
2 this. I still disagree.

3 DR. LIPICKY: I sort of got lost in the part where
4 you threw in Bob's presentation where this putative
5 difference was being displayed. I must admit, I didn't
6 quite follow it. But the only way one knows there is a
7 treatment effect ever is with respect to the historical
8 placebo. If the treatment circumstance is changing and one
9 is worried about imposing the historical placebo on the new
10 data set because it may not be effective -- is that the
11 question? So the historical placebo seems like a reasonable
12 thing to do, otherwise you don't really know it is even an
13 active control. So, that is thought on.

14 Thought two is that the relevant difference comes
15 to how different can the new drug be from the active, and
16 that has to be based on some guess with respect to what the
17 magnitude of treatment effect is because if the magnitude of
18 treatment effect is fairly large, you can have a fairly big
19 treatment effect from the new agent even if the point
20 estimate is less than the point estimate for the active. It
21 is still a pretty big effect. But if the treatment effect
22 is very small, then even you are just thinking point
23 estimates, the point estimate, if it is less than the active
24 control, may be totally of no benefit at all. So, it seems

1 to me, there has to be an estimate of the magnitude of
2 treatment effect. It has to come from somewhere. If one
3 says that is not stable, then one has no framework of
4 reference at all.

5 Then the second is that the decision of where the
6 new treatment has to position itself in relationship to the
7 active treatment needs to be in some confidence limit sense.
8 It doesn't seem to me that one can say on any basis that it
9 can be some fraction because of the point estimate because
10 you have the confidence limit problem. If you can make all
11 of those assertions that you can interpret the magnitude of
12 treatment effect, and you have some feeling for the
13 confidence limits and things like that, it seems to me you
14 must then accept the historical control or, if you throw
15 that away, you have no basis for comparing anything to
16 anything except for superiority.

17 DR. PACKER: But, Ray, as I understand it, even if
18 you were to be superior but your active control had very
19 wide confidence intervals you might not be able to reach
20 conclusions about its efficacy compared to a putative
21 placebo. Do you agree with that?

22 DR. LIPICKY: Well, I am not sure, but Dr. Temple
23 does not. He has his hand up.

24 DR. TEMPLE: Well, if you show superiority to an

1 active control the only thing you have to be sure of is that
2 the active control is not worse than placebo. And that is a
3 historical observation but you can be fairly confident in a
4 lot of areas that the active control is not worse than
5 placebo, then if you are superior to it, the results of that
6 trial are perfectly well interpretable. It is equivalent to
7 a placebo-control trial where you show a difference between
8 treatments. That is easy.

9 Let me give a couple of examples for some of the
10 other cases. Milton, think about carvedilol. You have
11 reviewed all the trials of ACE inhibitors against placebo in
12 looking at symptomatic change. What you found was that
13 about half of them, or something like that, were able to
14 distinguish drug from placebo. So, the historically
15 evaluated regular difference in trials that seem to be of
16 adequate size and design was that half of them couldn't tell
17 drug from placebo.

18 What that means is that, historically speaking, if
19 you now want to do a comparative trial of some drug with an
20 ACE inhibitor for symptomatic treatment of congestive heart
21 failure and try to say what is the guaranteed difference
22 between the active control and placebo that I will use and
23 what difference between test drug and my control group could
24 I describe that would show that the effect had been

1 preserved or lost, the answer would be there is no
2 difference that is regularly distinguishable by the control
3 drug. So, the only interpretable study would be one where
4 you would beat the control because you know historically
5 that many, many studies cannot distinguish drug from placebo
6 in congestive heart failure.

7 Another example that has been through this
8 Committee is in thrombolysis where a review of available
9 studies, probably five or six of them at the time, showed
10 that there was always at least a 20% or so benefit of
11 thrombolysis compared to placebo. So people concluded and,
12 in fact, this Committee in a previous iteration concluded
13 that you could reliably say that there was a difference
14 between the control drug and placebo, and you could identify
15 it. You could say it was about a 2% increase in mortality.
16 So if you then compared the new thrombolytic with a standard
17 thrombolytic, you could then ask what the difference is
18 between mortality of those two drugs. If the difference was
19 guaranteed to be less than 2% you could say with some
20 assurance, I have some effect, and more than placebo. But
21 the Committee didn't think that was good enough. They said
22 losing most of the effect of a thrombolytic is not
23 desirable. I want to preserve some guaranteed fraction of
24 that effect. So, at various times it said, I want to be

1 sure that the difference between these two treatments is
2 less than 1%. That would mean I am preserving at least 50%
3 of the historically derived effect. Some people said, no,
4 no, that's not good enough either. I want to preserve at
5 least 75% of the effect. In that case, the difference
6 between the two treatments would have to be less than 0.5%,
7 or the confidence interval would have to exclude a
8 difference greater than that.

9 That is what an equivalence trial turns out to be.
10 There is no such thing as an equivalence trial. All you can
11 do is say the difference is not larger than some amount.
12 You could use Bob's terminology equally well because the
13 guaranteed difference you can always detect is the putative-
14 placebo effect of the comparator drug compared to placebo.

15 DR. PACKER: It seems as if the distinction
16 between what you are saying and what Bob Fenichel is saying
17 can perhaps, in non-statistical terms, be summarized by what
18 you believe to be the truth about your active control or
19 what you can measure as the confidence intervals of that
20 truth. Is that a correct statement?

21 DR. TEMPLE: I am not sure whether we actually
22 disagree on this or not. Bob gave as the historically
23 derived drug effect -- he drew a figure and a confidence
24 interval around it. My problem is I think that trials

1 differ for a variety of reasons that we don't understand,
2 sometimes because of the population or whatever. In
3 depression, for example, what you see is that some trials
4 just don't show even a lean, even though the drug involved
5 is a well-established drug. You can either think that that
6 is a matter of variability or you can think that some
7 populations are no good at detecting things.

8 If you think the latter, then you really can't
9 have an active control equivalence trial or a putative-
10 placebo trial because there isn't any value that you could
11 attribute to the control drug. So, I have a qualitative
12 component to my putative placebo in addition to looking at
13 standard deviations and things like that. The beta-blocker
14 trials are an example of that.

15 DR. PACKER: Let me must try to explore the
16 difference that you just mentioned in just a little bit of
17 detail. The ACE inhibitor in heart failure issue is a good
18 example. As Ray said earlier, it is interesting to try to
19 define the standards of what is an active control. Clearly,
20 FDA approval of a trial may or may not be an adequate
21 standard in either direction. One good example would be ACE
22 inhibitors on exercise tolerance in heart failure. There
23 are many ACE inhibitors approved to enhance exercise
24 tolerance in heart failure but, Rob, as you mentioned, they

1 do so, so inconsistently that the confidence intervals of
2 such an effect are very, very wide.

3 DR. TEMPLE: Which means that in any given trial
4 where you compare a new drug with a control drug you cannot,
5 with assurance, say that the control drug would have beaten
6 the placebo had one been there. So, equivalence or lack of
7 a difference, or whatever you say, is just uninformative
8 because you don't know whether this is the sort of trial
9 that could have been informative about the difference
10 between the active drug and placebo.

11 DR. PACKER: I understand. That leads to one or
12 two questions that would imply that it is non-informative
13 for a company to ever try to show equivalence to an ACE
14 inhibitor in exercise tolerance.

15 DR. TEMPLE: That is certainly what we would tell
16 people.

17 DR. PACKER: So, if a company had an ACE inhibitor
18 that was approved for twice a day use and wanted to get the
19 Agency to change it once a day usage, and did a trial of
20 1000 patients, significantly larger than most exercise
21 trials, and showed that once a day and twice a day had equal
22 exercise capacity, that experiment almost invariably, in
23 that example, would be futile.

24 DR. TEMPLE: That is what I would say and that is

1 probably what Bob would say.

2 DR. FENICHEL: No, there is a difference in
3 approach that Dr. Temple has to my approach, and I think
4 maybe a better way to characterize the difference as the
5 center of our thinking about this is the placebo, which I
6 presume is constant in some sense. So, the active control
7 is really just a tool to get to the placebo. Perhaps this
8 comes out best if one considers a possible presentation,
9 which we have not had, which is suppose a new thrombolytic
10 came to us with two trials, one against streptokinase, say,
11 and one against TPA, and in both cases it was around the
12 same sense as the control drug and, because of perhaps the
13 sample size or because of variability or because of
14 something else, it really preserved a pretty good fraction
15 of the effect of TPA and not so good a fraction of the
16 effect of streptokinase. Some people would regard that as
17 paradoxical because they think TPA is a lot better than
18 streptokinase, others, including our next speaker, will say
19 it is not paradoxical at all necessarily because they are
20 really the same, and I don't mean to enter into that. So it
21 might be very difficult to describe this effect as
22 preserving fractions of something because it is a different
23 thing here and a different thing there. But one might also
24 describe that set of trials as saying, assuming this were

1 true, both times you have shown with such-and-such
2 confidence in one and such-and-such confidence in the other
3 that you are better than placebo.

4 There has been a descriptive matter of how good it
5 is? How confident are we, and what is the point estimate of
6 the effect size? And that might be very complicated. It
7 might have to do with unknown differences in subgroup
8 efficacy of the two different controls. There are all sorts
9 of possibilities. But it seems easier to describe the
10 result than to come to a regulatory conclusion about the
11 result if the fixed point is that of placebo.

12 DR. LIPICKY: I want to just return a little bit
13 to the discussion a little while ago, and that is it is not
14 clear to me that the description of an adequate active
15 control is where in every trial the active control can be
16 differentiated from placebo. To stick my neck out, the
17 exercise tolerance in CHF and the fact that from trial to
18 trial there is no reproducible winning against placebo only
19 says in trials of that size you cannot expect to
20 reproducibly beat placebo. So it could be that in larger
21 trials it would be a totally reproducible effect. It is
22 small and there is large variability. So, if one knew what
23 the reasonable point estimate of the treatment effect is
24 against placebo and what the variance of that is, then one

1 could decide what kind of sample size would be necessary in
2 the active control to be able to draw conclusions, and it
3 might be prohibitive.

4 DR. TEMPLE: That is only true if what you are
5 looking at is something where there is no what you could
6 call study by treatment interaction, and I don't think we
7 necessarily know whether there is. In depression, I would
8 allege, some populations just don't respond and you don't
9 know the reason. If someone could define a study sample
10 size or a study population in which you could always win,
11 then that would be okay even if some other sample sizes and
12 other studies didn't. But the burden is on someone wanting
13 to use this design to make that case.

14 DR. LIPICKY: Right, but it does not have to be
15 that every trial that has ever been done, or 90% of the
16 trials that have ever been done or 70% of the trials that
17 have ever been done, have to have demonstrated a
18 superiority.

19 DR. TEMPLE: Not if you know the reason for
20 failure.

21 DR. LIPICKY: Right.

22 DR. THADANI: I think one of the difficulties I am
23 having is the moving target of the active controls whether
24 you are using not the historical but even placebo. But with

1 changing time the background treatment could have changed.
2 Take, for example, antianginals. We take it for granted
3 beta-blockers and ACE inhibitors work but suppose you had a
4 population in which the majority of the patients already
5 have had bypass surgery, we have no idea whatsoever how the
6 response rate of those patients is because a lot of patients
7 did not enter the studies in previous trials. That could
8 affect all your results. You know, it is not a mortality
9 trial but I think it depends on what you are looking at.
10 So, in the past we required even two trials to go in the
11 same direction or at least to give us some confidence. So,
12 unless one can exactly define the population which entered
13 the previous trials and do the next trial with a very
14 similar population, then I think the conclusions might be
15 very wrong, and maybe that is the reason we are having
16 different answers. Even when you look at meta-analysis of
17 even aspirin or whatever, different populations went in and
18 it makes it very difficult -- or thrombolytics looking at
19 mortality, looking at large enough trials. But if the
20 trials are small, even with the confidence intervals I am
21 not sure, as you said, if the next trial might go the wrong
22 way. So I am confused on that issue.

23 DR. D'AGOSTINO: I have a couple of comments but I
24 think that the notion of what else exists out there is very

1 important in the sense of after you have done your
2 equivalency trial, or whatever you want to call it, you do
3 have to put it in the context of what we have out there. In
4 analgesic trials, for example, a lot of them even against
5 placebo, and certainly against actives, don't come out to
6 show anything, and I am not sure you can trace that down and
7 say let me explain this trial; let me explain that trial and
8 I will understand the population where it works and doesn't
9 work. I am not sure we are that clever. So, I think if
10 there is a lot of history out there that says that active
11 control trials are going to be problematic, we start off on
12 a very, very bad footing saying that we are going to do an
13 active and then make a comparison with some placebo that we
14 think we might know.

15 The other thing is that I guess I get lost in all
16 the vocabulary that the statisticians have generated. I
17 don't know what is wrong with them. But aren't we basically
18 trying to show at the end that after we have something from
19 our active control trial we wish we had a placebo and we
20 want to make a comparison with the placebo? Isn't that the
21 basis of it? I mean, we get carried away with all the
22 vocabulary but isn't that what we are doing? I mean, there
23 may be many, many ways of doing that but this two times the
24 sigma, four times the sigma --

1 DR. TEMPLE: That is what everybody is saying in
2 one way or another. The way the international document is
3 coming out, you define a margin that is the entire
4 difference between the placebo, had there been one, and the
5 control drug. And if you can't be sure that you haven't
6 excluded a difference greater than that, you lose.

7 You are right, in many cases, like analgesics, you
8 could never describe such a difference because many trials
9 fail. Depression would be the same; anxiety would be the
10 same; angina would be the same; heart failure would be the
11 same.

12 DR. D'AGOSTINO: What I am concerned about is we
13 get caught up in the discussion with the word sigma, where
14 that is one way of attacking it, which I hope we aren't
15 locked into. I think that is a way that one can approach
16 the problem but we are spending more time trying to
17 understand what the four sigma is saying than we are --

18 DR. TEMPLE: In some ways for better or worse, and
19 maybe this is because there have been clinicians involved in
20 it, there has been a tendency to set the margin, the
21 difference that you have to not be greater than,
22 irrespective of confidence intervals; just to pick a number
23 and then say I want to be sure, two standard deviations
24 worth, that I am not worse than that. So, it is actually

1 conventional difference testing and analysis, and the bound
2 of the confidence interval has to exclude a difference
3 between the treatments greater than that because if it is
4 greater than that you have lost all the effect you thought
5 you had. But it is very much what Bob was showing. It
6 looks very much the same.

7 DR. CALIFF: I guess I am dense but I don't think
8 the question, if you have a treatment which is already
9 effective, is how would the new treatment compare to
10 placebo? I mean, that is irrelevant. What the patient
11 needs to know is, is the new treatment within some
12 reasonable range of what the standard treatment provides.
13 Bob, you said one key thing which is different than just
14 having to show you are better than placebo. When you gave
15 your example you said and within a minimally important
16 difference which is tolerable to the clinical situation that
17 I am certain you would be willing to accept about how much
18 worse it could be. To me, that is a very different thing
19 from saying that what you really want to know is what a
20 placebo would have done.

21 DR. D'AGOSTINO: That is the question I was
22 asking. Is the regulatory thing saying that you beat the
23 placebo, or are you trying to say that I now have an active
24 control --

1 DR. CALIFF: My concern about this is that the
2 regulatory thing says you just have to beat the placebo, and
3 we are going to have a bunch of therapies out there which
4 are better than placebo but maybe worse than the current
5 standard and there will be no motivation to answer that
6 question.

7 DR. TEMPLE: No.

8 DR. LIPICKY: No, that is not true.

9 DR. TEMPLE: That is a different question. The
10 first question, Rob, that we struggled with is, I mean, the
11 usual test for whether a drug works is a comparison with
12 placebo. You usually test at the 0.05, which means that the
13 lower bound of your confidence interval is just above no
14 effect at all, and that is usually considered acceptable.
15 Now, you know, the point estimate is really higher than that
16 so it is not very likely that it is minimally effective. It
17 is much more likely that it has some measurable effect.

18 If you now have a drug that is a pain medication,
19 for example, you can say, well, I should apply the same test
20 as I always do: I want this thing to be better than nothing.
21 That is the test for an analgesic usually. The equivalence
22 to showing that something is better than placebo, in active
23 control terms, is that I am positive I have preserved some
24 of the effect of the active control -- some of it. If you

1 are satisfied with a drug that is better than nothing, then
2 that is an appropriate proof standard. What you are saying
3 is that sometimes when there is a treatment in the community
4 that we know to be valuable, we want more assurance than
5 that. We want to know that some fraction of it is
6 preserved. That is the thrombolytic example. Because that
7 is a mortality effect, most people --

8 DR. CALIFF: Or the ACE inhibitor effect.

9 DR. TEMPLE: Or an ACE inhibitor. Of course,
10 given what you know about the results of ACE inhibitors and
11 heart failure in trials in symptomatic disease, confidence
12 that it is better than placebo is about what you have for
13 the available drugs. If you are now talking about the
14 survival effects of ACE inhibitors you might say, no, I want
15 to preserve at least "blank" percent of it. But that is a
16 separate clinical judgment that you impose. The
17 mathematical thinking is the same --

18 DR. CALIFF: I agree, the thinking is the same but
19 the question is whether the regulatory standard is beating
20 placebo or the regulatory standard --

21 DR. TEMPLE: Well, that is what we have advisory
22 committees for.

23 DR. CALIFF: Well, I want to argue for more than
24 just beating placebo for the reasons I have articulated.

1 The second brief point is I just don't buy this
2 magical thinking that somehow drugs work and sometimes they
3 don't work at other times. I think when you try to do
4 studies that have a minimally acceptable sample size you are
5 going to hit and miss. I think I really agree with what Ray
6 said. If adequate size studies were done you would get true
7 effect.

8 DR. TEMPLE: Why do you need a bigger study one
9 time than another time?

10 DR. CALIFF: Excuse me?

11 DR. TEMPLE: Why do you need a bigger study one
12 time than another time?

13 DR. CALIFF: Because there is variance in
14 responses in different populations.

15 DR. TEMPLE: There is variance in response in
16 different populations?

17 DR. CALIFF: Yes.

18 DR. TEMPLE: Well, that is the same as saying
19 there is a study by treatment interaction. It is no
20 different. When I say sometimes antidepressants can't show
21 any effect you want to say, well, it takes a bigger study
22 one time to show an effect than another time. We are saying
23 the same thing. It means that you can't define ahead of
24 time, unless you do it, a study of a certain size, of a

1 certain design, that can regularly distinguish drug from
2 placebo. If you can do it, be my guest.

3 DR. CALIFF: Well, I think it can be done, and I
4 think the reason it is not done is people try to do the
5 minimal sample size and it has never really been looked at.

6 DR. TEMPLE: In a certain sense we don't care. As
7 soon as someone shows a study of a particular size and
8 particular design that can regularly distinguish drug from
9 placebo, you are in the active control business. But until
10 you do that, and nobody has done it for analgesics or heart
11 failure, obviously, or in symptomatic heart failure, or
12 angina, or depression, or anxiety, or all of those things,
13 then you can't use the model Bob is talking about because
14 you can't identify a guaranteed difference between the drug
15 and placebo.

16 From looking at depressing trials, I think if you
17 looked at them all you would say there really is a
18 difference in populations, and that some populations either
19 don't respond or respond.

20 DR. LIPICKY: Well, maybe we need to do that in
21 order to resolve this. Has anybody actually done it?

22 DR. TEMPLE: You can't, Ray. Somebody has to do
23 huge trials. Why should they bother?

24 DR. LIPICKY: Well, I have heard of meta-analyses.

1 DR. TEMPLE: Meta analysis doesn't help you.

2 DR. CALIFF: Now we have a bunch of
3 antidepressants out there and we have no idea how one
4 compares to the other, or what the long-term health effects
5 are, or how they deal with general populations.

6 DR. TEMPLE: You actually know more than you
7 think. There have been thousands of comparisons and they
8 never managed to show a difference. So, the answer is they
9 are probably all about the same.

10 DR. PACKER: Rob, you said one thing I just want
11 to clarify. You have emphasized that a lot of the level of
12 uncertainty is due to inadequate sample size. I just want
13 to make the point that adequate sample size is not
14 necessarily always a solution. For example, in the
15 situation with ACE inhibitors there is a reason to believe
16 that the variability in exercise tolerance would increase as
17 the sample size increased, so that your confidence intervals
18 would not necessarily become narrow if you went from a
19 clinical trial of exercise from 300 to 1000 because there is
20 tremendous variability in exercise performance from center
21 to center. It is that kind of endpoint.

22 DR. CALIFF: It just means you are measuring a
23 worthless endpoint.

24 DR. PACKER: Well, that may be true, but it is an

1 endpoint which one could use in an active controlled trial.

2 DR. CALIFF: But probably shouldn't if that is the
3 property of the endpoint.

4 DR. PACKER: I think that is a good point.

5 DR. LIPICKY: So, you think it is necessary to
6 always use as an active control an agent that every trial
7 always distinguishes it from placebo. Is that where this is
8 leading? I think that is dead wrong, and maybe we need to
9 lay that out somehow. You may be right and your intuition
10 may be the correct intuition, but my intuition leads me in
11 different directions, and maybe we should lay that out
12 sometime because that is an issue -- how would one pick the
13 active control? But that is only an issue -- right? -- as I
14 see it, if one thinks one needs to have an estimate of the
15 magnitude of treatment effect. If you think you can do
16 without that number and then evaluate a positive control
17 trial, then the discussion that has just been going on is of
18 no consequence.

19 But that raises the second question that I wanted
20 to ask and I am wondering where that sits, that is, if the
21 notion is that one doesn't need to know the magnitude of
22 treatment effect and/or its variance for purposes of
23 evaluating a positive control trial, then one doesn't have
24 to rely on historical controls. Then one would say, well,

1 you shouldn't rely on historical controls because the
2 treatment circumstances have changed over the course of time
3 and the magnitude of this treatment effect may have gone
4 away or be very different. But if that is true and one can
5 make an argument that that is true, then it seems to me it
6 also follows that you don't know the active control works
7 any more. And if one can make that argument persuasively,
8 you can do a placebo-controlled trial, and you do not need a
9 positive controlled trial because there is no argument that
10 you know the active drug works.

11 So, it seems to me these are logically
12 contradictory things to be saying, and I am not quite
13 following the arguments.

14 DR. CALIFF: Wait a minute, wait a minute. The
15 argument is not that you can't -- well, first of all, we
16 would all agree that we have a level of uncertainty as time
17 passes and new therapies are introduced whether the old
18 therapies have the same effect they had before, and there is
19 danger either way. There is no right, secure answer. You
20 have to take some risk either way. But I would argue pretty
21 strongly that I would be willing to take a risk that the
22 treatment that was shown to be superior to placebo in
23 definitive trials probably still is in the future. That is
24 a risk worth taking. Assuming that we know the magnitude of

1 that effect ten years later when there are eight other
2 effective treatments that all these patients are getting, I
3 think is --

4 DR. LIPICKY: But then you are arguing you don't
5 need to know the magnitude of the treatment effect in order
6 to evaluate a positive control, and I don't see how you can
7 do that.

8 DR. CALIFF: I am arguing that there is a risk
9 either way, but I would prefer to take the risk on the side
10 of comparison --

11 DR. LIPICKY: But how do you propose to evaluate a
12 positive control trial without having any insight into what
13 the magnitude of the treatment effect of the positive
14 control is? It seems to me you have to have that number
15 somehow, otherwise you are at sea.

16 DR. CALIFF: You either have to guess what you
17 think the putative placebo would be doing over time relative
18 to the active control, or you have to say we have a standard
19 treatment and we are comparing a classical "equivalence"
20 design.

21 DR. LIPICKY: But the first example that you gave
22 would be to say I am going to throw away the historical
23 control data, and my guess is better. I suppose you could
24 try and defend that but it would be hard to.

1 DR. TEMPLE: They are the same thing. Your
2 historical control estimate is your best guess. If you want
3 to say, well, I think over time the difference has probably
4 shrunk because the background rate has declined, then you
5 build that into the difference that you are trying to
6 exclude.

7 Rob, I don't understand what distinction you are
8 making. You can't interpret an active control equivalence
9 type trial without making some estimate as to what the
10 effect of the control is versus placebo because if you don't
11 do that, you don't know what kind of difference between
12 treatments you have pulled out. There is no such thing as
13 equivalence. All you can ever say is the difference is not
14 greater than thus-and-such. That is all you can ever say,

15 DR. CALIFF: Right.

16 DR. TEMPLE: The thus-and-such is the numb of all
17 this. The only way you can define it in an active control
18 trial with no placebo is historically. You can't escape
19 that burden; you have to do it.

20 DR. CALIFF: I don't object to going through the
21 exercise. I just want to make sure the definition is not
22 beating placebo; the definition is coming within a
23 clinically relevant difference from the active control.

24 DR. TEMPLE: Let's take something where there is

1 no mortality effect so we are not worrying about that. It
2 is a pain pill, symptomatic treatment. The current standard
3 for approval now is you beat placebo; you show you are
4 better than nothing. We tend to believe the point estimates
5 even though we probably shouldn't. That is just the
6 standard. You beat it at 0.05 or something like that. That
7 doesn't have to be. You can say I have to have an effect of
8 at least this and you can make an effect smaller than that
9 your null hypothesis. We don't have to say better than
10 nothing is sufficient, but we historically do and
11 historically, by the way, it is not that easy to do that in
12 many drug classes. It is hard to beat placebo.

13 So, I guess I would put to you in a symptomatic
14 treatment, in an active control, if you can be sure that you
15 are better than nothing and the point estimates are roughly
16 in the right place, you have done what you usually do.
17 Which standard would you then impose? We could make the
18 standard higher.

19 DR. CALIFF: Well, the conclusion from your trial,
20 when I am the patient in the dentist's chair and I have pain
21 control method A and B, and pain control method A is
22 something that has been around for a while and we know how
23 it works and what its general effects are, and now we have
24 the new treatment, B, and my dentist says treatment B is

1 better than placebo but I have no earthly idea really how it
2 compares to A, and we define whether or not we give you
3 treatment B by how it compared to placebo and not by how it
4 compared to A, the question I would have is which one is
5 better.

6 DR. TEMPLE: Suppose now you are talking in an
7 area where you can still do placebo-control trials, are you
8 saying that there should always, in addition, be an active
9 control trial because we shouldn't approve drugs unless they
10 are better than, or some fraction of the available therapy?
11 You have a legal problem if you say that.

12 DR. PACKER: Let's just put a bookmark here for a
13 moment because there are other chapters this morning. Let
14 me quickly ask Marv, JoAnn and Dan for a brief comment,
15 hopefully brief comment, because we have to go on with the
16 rest of the program.

17 DR. LINDENFELD: Just briefly, I was concerned
18 about what was brought up earlier, that the placebo group
19 versus active control may in some cases contain four new
20 medications, and how would you know the magnitude effect
21 historically? That is a very difficult point.

22 DR. KONSTAM: I just want to say I think most of
23 this discussion seems to me to relate to what the
24 appropriate methodologies are to reach a philosophical

1 conclusion about what appropriate regulation is. I must
2 say, I hear Rob saying something different. I hear Rob, and
3 I don't want to go back and forth but I hear Rob challenging
4 what I am taking to be the basic regulatory principle that
5 we seek to determine whether a drug is different from
6 placebo. I hear Rob saying, no, it is not really the basic
7 philosophical underpinning that we should be striving for.
8 We should be striving for improving the clinical
9 opportunities out there. As a clinician, I sympathize with
10 that but, for me, I would like to hear a clear philosophical
11 statement that the ultimate goal is to say this drug does
12 something; this drug works better than placebo. That is
13 equivalent to saying this drug does something positive.
14 Maybe it is not practical that you would ever see this, but
15 if you were convinced that the drug is better than placebo,
16 although slightly worse than other available therapies, and
17 if you could know that, would you approve it or not? Based
18 on the construct that I, and I think others, have been going
19 on, I would say, yes, I would approve it and maybe there are
20 circumstances where it would be used. And there are a lot
21 of other issues. But I think, for me, I am going to need
22 some kind of clarification about the basic philosophical
23 construct that we are under in terms of that.

24 DR. RODEN: I think my comment is the same as

1 Marv's. Basically, pretend we have a new thrombolytic which
2 is in an active control trial and is demonstrably 75% as
3 good as standard therapies, is that a basis for approval
4 because it is better than placebo? I don't think I want an
5 answer to that right now, but I am not sure I would agree
6 with Marv that that is a basis for approval.

7 DR. PACKER: We will get some more clarification
8 on these issues in a few minutes. We will go on with the
9 next speaker, Dr. Rory Collins. We are glad to have him
10 with us, having traveled quite a distance to participate in
11 today's meeting. I guess the title, Rory, of your
12 presentation is "If That is Your View, Then This is What You
13 Have to Think About." He is going to discuss something like
14 that.

15 **If That is Your View, Then This is What**
16 **You Have to Think About**

17 DR. COLLINS: Thanks very much for the opportunity
18 to come and talk, and it was nice to have the general
19 discussion earlier than anticipated because it at least
20 encouraged me to realize that it wasn't just me that was
21 confused about the issue.

22 I think that the purpose of equivalence trials is
23 actually not to demonstrate equivalence, and I must say, I
24 find great sympathy with what Califf is saying in that I

1 think the intention is to determine not equivalence, but to
2 demonstrate that the new treatment is effective, and to get
3 some idea of how effective. It may still be that one would
4 want to use a treatment that was less effective than a
5 standard treatment because there are cost advantages,
6 convenience advantages, or whatever. But you would actually
7 want to know how effective a treatment was, and you would
8 want to know that it was effective. I think those are the
9 aims.

10 (Slide)

11 I think most of what I am going to say is, I hope,
12 self-evident but the reasons for positive control trials,
13 certainly the reasons that have been given, are that there
14 is a standard treatment with proven efficacy so for some
15 reason a no treatment comparison group is considered
16 inappropriate.

17 The new treatment is expected to have similar
18 efficacy, or maybe greater efficacy. If it has similar
19 efficacy, then it still might be of interest because of
20 safety or convenience or cost advantages. I mean, it is
21 very clear that this has often led to a number of direct
22 comparisons of treatment B versus treatment A in what I am
23 going to term positive controlled trials or active
24 controlled trials.

1 But I would just like to take a minute to say that
2 if one can get away from the sort of confusion that there is
3 around equivalence trials, if one can get away from doing
4 positive control trials, then wherever one can one should do
5 so. I would just like to encourage the greater use of "add-
6 on" studies, and this is certainly something that the FDA,
7 and Bob Temple has written on. I am saying that it may be
8 possible and more appropriate to do an add-on study of
9 treatment B plus treatment A versus the same treatment A.

10 (Slide)

11 There are a lot of advantages of doing that. When
12 might one do an add-on study? Well, obviously if there is
13 still an increased risk of the adverse outcome even with the
14 standard treatment so that risk that you would like to
15 reduce; if the new therapy that you are thinking about
16 produces its effects at least largely through a different
17 mechanism, or at least you believe it does; and if the
18 combination is reasonably well tolerated, then in those
19 circumstances it would have to be much better if your aim is
20 to determine that the treatments are effective. It would be
21 much better to do an add-on study because the difference
22 between treatment and no treatment is likely to be bigger
23 than the difference between two active treatments. So the
24 difference between B plus A versus nil plus A, which is

1 essentially active versus no active, is likely to be bigger
2 than a direct active comparison.

3 (Slide)

4 So, it ought to be easier to demonstrate that the
5 new treatment is effective. I will just take one example.
6 This is an example of blood pressure. There is continuous
7 relationship between blood pressure and stroke, well down to
8 the levels of blood pressure that is far below the target
9 levels in guidelines. The available treatments that we have
10 generally produce relatively modest reductions in blood
11 pressure but by combination therapy that is used still the
12 targets that are achieved lower blood pressure, at least
13 epidemiologically, would be expected to be at lower risk.
14 And the combinations in general would be well tolerated, at
15 least in most patients.

16 Despite this, most of the large-scale
17 antihypertensive comparisons that have been going on go to
18 great effort to try to achieve similar blood pressure levels
19 in the two treatment groups. I mean, it is crazy;
20 completely nuts. If lower blood pressure presumably would
21 lower risk, at least that is what epidemiology would
22 suppose, add-on comparisons would actually be clinically and
23 scientifically much more appropriate and, as I mentioned
24 before, the difference between treatment and no treatment,

1 the difference between greater blood pressure reduction and
2 lower blood pressure reduction should be bigger than the
3 difference between two equivalent blood pressure levels.

4 (Slide)

5 Here is one example where add-on comparisons would
6 be much more appropriate and, yet, they are not done. But,
7 I mean, you could think of more. In breast cancer we have
8 all been comparing the acute cytotoxic chemotherapy with
9 hormonal therapy. Add-on studies would be better of
10 chemotherapy plus tamoxifen versus tamoxifen, for example.
11 Anti-platelet therapy to prevent vascular events. You
12 wouldn't want to really compare, say, aspirin versus
13 dipyridamole, two agents working through different
14 mechanisms. Combination of the two versus one alone is a
15 much better approach. Or, anticoagulants versus anti-
16 platelet therapy is a silly comparison as a direct
17 comparison.

18 I have mentioned blood pressure lowering but,
19 look, we are bound to get a new cholesterol-lowering class
20 of drugs. People are working on them. So, do we want to
21 compare them with statin? No. We believe that lower
22 cholesterol would produce lower risk. Add-on studies would
23 be much better.

24 (Slide)

1 I think we really need to work at avoiding
2 positive control studies. So my feeling would be for not
3 positive or null control studies wherever possible. If you
4 have to do a positive control study, then think about doing
5 a combination or positive and null control because here you
6 could look to see whether it is effective in the presence of
7 A, and whether A is effective in the presence of B, as well
8 as having a direct comparison.

9 (Slide)

10 So, coming back from diversion and my plea for
11 avoiding positive control studies, and back to positive
12 control studies, obviously there are two different types.
13 There is the positive control superiority trial where you
14 are aiming to demonstrate that something is better and,
15 essentially, that is like a null control study
16 methodologically. There is no particular difference
17 philosophically in the approach. It is just less
18 interesting than add-on studies because, as I said before,
19 the difference between two active agents is likely to be
20 smaller. So if you have something that is superior, think
21 about would it be even more superior if you added it to
22 standard?

23 So, coming down to the positive control
24 equivalence studies, the aim there is to demonstrate lack of

1 any worthwhile difference in outcome between the two active
2 treatments. But really the point is then to say indirectly
3 that each of the treatments is better than no treatment.
4 And these are very different methodologically. I am going
5 to try to avoid methodology because Dr. DeMets is going to
6 talk about that in detail. But these two things are
7 diametrically opposite from each other.

8 (Slide)

9 This is my hypothesis -- a null control or
10 positive control superiority trial, the null hypothesis that
11 you beat is about the same as zero in the null control, or B
12 is about the same as A in the superiority trial. So, they
13 are philosophically the same.

14 But a positive control equivalence trial is
15 totally the reverse. The null hypothesis is that the effect
16 of B is not equal to A, and the alternative, the effect of B
17 equals A.

18 There is lots of sloppy writing, I think, in the
19 interpretation of trials. If you have not rejected the null
20 hypothesis, then failure to reject the null hypothesis in a
21 null control or positive control superiority trial does not
22 imply equivalence. The lack of evidence of difference is
23 not the same as evidence of lack of difference. Similarly,
24 in a positive control equivalence trial, failure to reject

1 the null hypothesis, that is, you are concluding that B is
2 not equal to A -- you can't really conclude that. It may
3 well be that they are equivalent. You just haven't been
4 able to reject the null hypothesis. So, they are reverse,
5 and this causes a lot of problems philosophically, as we saw
6 earlier.

7 (Slide)

8 So again, the advantages of null control trials --
9 well, the difference in outcome between treatment and no
10 treatment is likely to be larger. The appropriate design
11 and conduct of trials, null control and positive control
12 superiority trials, reduce the likelihood of falsely
13 concluding that there is no difference in outcome. That is
14 not true in equivalence studies, or not necessarily true.

15 Now, standard intention-to-treat analyses where
16 one compares all of those randomized to one treatment group
17 compared to all of those randomized to another group tend to
18 be conservative in such studies, in null control studies.
19 So they tend to diminish apparent differences between the
20 treatment groups. So, of course, in an equivalence trial
21 that may result in falsely concluding that there is
22 equivalence. So, it is the opposite. It is not
23 conservative. And rejection of the null hypothesis in null
24 control studies implies not only that a difference exists

1 but also that the trial was competent to detect it. Stephen
2 Senn, from London, has pointed out that in an equivalence
3 study the only time you can be absolutely sure that the
4 equivalence trial was competent was when it actually rejects
5 the claim of equivalence. So, the only time you know it is
6 competent is when there isn't equivalence, which is not
7 terribly helpful.

8 (Slide)

9 What about estimation in effects of treatment? I
10 think that is a key point that particularly Dr. Califf was
11 mentioning during the discussion. We want to know not only
12 if a treatment is effective, but how effective it is. Well,
13 it has been discussed that the control in a positive control
14 study may differ in the new trial from the effects in the
15 previous null control trials that demonstrated that the
16 standard worked. The differences in the patient population,
17 in a thrombolytic trial if you treated only patients within
18 six hours or within 24 hours you would get different
19 proportional effects, or high and low risk individuals --
20 differences in concomitant treatment.

21 I think this is a much bigger problem when one
22 looks at differences in absolute risk or in absolute risks
23 in a trial. They may well differ very substantially in
24 different circumstances. It may well be that it is better

1 to look at the proportional effects of treatment on risk of
2 particular adverse outcomes, and these may be quite a lot
3 more similar in different circumstances. For example, the
4 proportional reductions in stroke with antihypertensive
5 therapy are very similar in primary prevention studies and
6 in secondary prevention studies, or in people with very high
7 blood pressure or very low blood pressure with the same
8 blood pressure reduction.

9 (Slide)

10 If you look at anti-platelet therapy, as another
11 example, comparing anti-platelet therapy versus nil -- this
12 is just looking at the effect of anti-platelet therapy
13 versus no anti-platelet therapy on major vascular events,
14 MI, stroke or vascular events in prior MI, acute MI, prior
15 stroke with TIA or high risk individuals, then in these
16 different settings the absolute risks in the control group
17 are quite different, 17%, 10%, 14%, 20%. The proportional
18 reductions in risk though are quite similar even though the
19 absolute risks and the absolute difference in risks are
20 different. So, perhaps proportional differences will be a
21 better way of combining the data from a null control and a
22 positive control study.

23 (Slide)

24 So we want to demonstrate efficacy by combining

1 the effects of positive control trials, new versus standard,
2 and a null control trial, the standard versus nil. I am
3 talking now about the inequivalence trials. I mean, if the
4 new is better than the standard then everything is simple
5 again.

6 We have to take account of the biological
7 variation between these different types of trials conducted
8 in different circumstances and different times, and I have
9 no idea how one does that, other than waving hands and just
10 being a little less certain about the results, and maybe one
11 could build that in, in the statistical analysis, having
12 wider confidence intervals and things like that. That
13 certainly would be an approach that I have taken in the
14 examples I will show.

15 It is important to take into account the
16 statistical variation in the results of both types of
17 trials. So, not only the variation in the assessment in the
18 new versus standard, but also in the assessment of the
19 effect of the standard treatment from the standard versus
20 nil.

21 (Slide)

22 That is quite often not done. Interestingly, if
23 you want to combine the proportional effect, then it is
24 actually very simple to do because you can just add up the

1 log odds ratios from the trial of the standard versus nil
2 and new versus standard. The log odds ratio as an estimate
3 of the new versus nil as an estimate of the efficacy of the
4 new treatment, even though you are not comparing it against
5 nil, can be obtained by looking just at the sum of the two
6 log odds ratios with variance which is equal to the sum of
7 the variances of the log odds ratio. So, mathematically --
8 I mean, you could do it lots of different ways but if one is
9 looking proportionally there is quite a simple way of doing
10 that. You can then use that to estimate the reduction in
11 risk and confidence intervals around that reduction in risk,
12 and I aim to use this in a couple of examples.

13 (Slide)

14 I don't think it is the statistics that is the
15 problem. The problem is what is the source of the estimated
16 effect of standard treatment. Is it one particular trial
17 whose results you like? Maybe it has a very extreme effect.
18 If you put in a very extreme estimate, then it is going to
19 be easier to demonstrate that your new treatment isn't as
20 bad as placebo. Or, is it an overview of the related
21 trials, even though those trials may involve a range of
22 different treatments? In fibrinolytic therapy, for example,
23 all the trials of fibrinolytic therapy versus controls? Or,
24 should you just take the trials of SK versus placebo? Or

1 even a subgroup of the trials? I mean, all the trials
2 including people early or late? And we do know that the
3 effects are small in people treated later. So, maybe a
4 subgroup. So, there is a lot of uncertainty about the
5 estimate of efficacy of the standard.

6 There is the difficulty that the similarity of
7 proportional effects may correspond to dissimilarity in
8 absolute effects, which, in the final analysis, is what we
9 are interested in. It is the absolute difference we are
10 interested in. The question is how to estimate it.

11 So, one very good way of making treatments look
12 similar is to test them in low risk individuals and compare
13 absolute differences. But also, if we are going to base our
14 estimates on similarity of proportional effects, that may
15 not translate into similarity in different circumstances.

16 The balance of a reduction in one type of event
17 and increase in another may differ in different
18 circumstances. So, whereas the combination may be
19 equivalent in one circumstance, if there is a small increase
20 in stroke, say, and a small decrease in mortality in the
21 setting of the trial, when you translate that into another
22 setting where maybe the background risk of stroke is much
23 higher then you may not have equivalence.

24 Then the final problem is how much of the

1 estimated advantage, and do you mean proportional or
2 absolute, of the standard treatment must be guaranteed for
3 the new treatment in order to conclude that they are
4 equivalent or perhaps better to conclude that the new
5 treatment is worth having?

6 (Slide)

7 Just to touch on composites, I think composite
8 outcomes can obscure lack of equivalence. So, if you want a
9 tip on how to make things look equivalent use composite
10 outcomes. If, for example, you have a trial of 10,000
11 versus 1000 stroke, 180 versus 120, excess of 6/1000, highly
12 significant non-stroke death, 700 versus 800, so a
13 significant reduction with the new treatment of 10 versus
14 10/1000, if you looked to the composite outcome you would
15 conclude, perhaps falsely, that there is equivalence. There
16 is a difference of 4/1000. Of course, it would depend on
17 which population you did this study in as to whether you
18 would get this balance, or if the new treatment looked
19 better, or the new treatment looked worse. So, it may be
20 much better, if one is interested in determining
21 equivalence, to look at outcomes separately, particularly
22 outcomes that might go in opposite directions rather than to
23 look at composites. And there have been suggestions of
24 adding on to the composites, like this, outcomes that

1 haven't been shown to be influenced by the standard
2 treatment which can, again, even further obscure differences
3 between treatments.

4 I have seen papers on equivalence of thrombolytic
5 therapy where recurrent angina has been included in the
6 composite. Well, there is no evidence that thrombolytic
7 affects that outcome anyway. So, it would make the
8 treatments look more equivalent.

9 (Slide)

10 So, a couple of examples that we touched on. I
11 just took some quotes out of the report of the INJECT study
12 which compared reteplase versus streptokinase. I think this
13 was just trying to summarize the thinking that was going on
14 in the design of that study.

15 ISIS-3 and GUSTO studies showed the size of study
16 needed to identify a difference in mortality of 1%. That is
17 very big. Equivalence trials offer an alternative. There
18 have been papers written by the group saying that
19 equivalence trials offer an alternative to mega trials; that
20 they can be smaller. I mean, the muddled thinking that is
21 going on is extraordinary. To determine equivalence will
22 require bigger trials, not much smaller trials. But this is
23 offered as an alternative.

24 Although this trial is an equivalence trial, its

1 rationale differs from that of a conventional equivalence
2 trial. The starting point was that they weren't equivalent.
3 It was the belief that reteplase offers a small mortality
4 benefit. I mean, it is a very interesting approach. Then a
5 new agent should be an acceptable alternative to a standard
6 agent if the mortality rate for the new agent is not more
7 than 1% worse than the standard. That has obviously been
8 plucked out of the air as an estimate of how much of the
9 putative effect of the thrombolytic therapy versus nil is
10 worth keeping. The conclusion was that reteplase is an
11 effective drug in the treatment of acute MI. It is at least
12 equivalent to streptokinase.

13 (Slide)

14 I wanted to look at the results and see whether
15 one could conclude that they are equivalent. I am going to
16 combine both odds ratios and show you the results.

17 Here is the direct comparison of fibrinolytic
18 therapy versus nil from a combination of the randomized
19 trials, looking at patients with ST segment elevation within
20 12 hours, which was the category that INJECT was thinking
21 about. So a 24% reduction with 99% confidence intervals,
22 going from about 17% to 30%. I have put in 99% because I
23 think one needs a little bit more uncertainty when thinking
24 about what are effectively historical comparisons.

1 So, if we then combine the results of INJECT with
2 this result to say what is the effect of rPA versus nil,
3 combining INJECT plus FTT plus the overview, then we get a
4 point estimate of 28%, but with a lower confidence interval
5 of about a 9% proportional reduction. So, let's say there
6 is a 10% absolute mortality, you are preventing 25 deaths
7 per 1000 patients. But it might only be 9/1000 with rPA --
8 maybe, being pessimistic.

9 The GUSTO-III study is a bigger study, I think
10 taking random error more seriously. So, you can see the
11 standard deviation is narrower. But, still, in that study
12 5% versus 24% or 25 versus 5/1000, 10% absolute mortality.
13 And maybe the best estimate for rPA is to combine INJECT
14 plus FTT and GUSTO plus FTT, and we can get an indirect
15 meta-analysis of the two trials to say what is the effect of
16 rPA versus nil, and we are moving the lower limit of the
17 confidence interval away from zero but whether one is
18 comfortable with the possibility of preventing only 10/1000
19 rather than something like even the lower limit of 17/1000
20 is debatable. I am certainly not going to come out with
21 solutions but I can try and describe the problem.

22 (Slide)

23 There was a very nice editorial just this last
24 week, from Elliott Antman, in the New England Journal

1 commenting on the COBALT and the GUSTO-III studies which
2 were direct comparisons of thrombolytic therapy. There were
3 some interesting comments particularly about the COBALT
4 study.

5 In that trial, he says, the calculation of the
6 sample size was based on the assumption that double-bolus
7 administration of TPA would actually reduce 30-day mortality
8 from 6.3%, which is what was seen in GUSTO with accelerated
9 TPA, to 5.4% based on surrogate outcomes of angiographic
10 data. As a result of assuming that they aren't equivalent,
11 if the true mortalities were identical, Elliott Antman said,
12 say, 7.5% in each group, which would seem to be a reasonable
13 way to calculate power calculations before you have a
14 result, then the probability of demonstrating equivalence by
15 the COBALT criteria, which was a difference of 0.4% in
16 absolute terms, was only 0.16. It has 16% power. You
17 wouldn't get that from reading the actual report.

18 An equivalence trial designed to rule out, with
19 80% power, excess mortality of 0.4% when the true mortality
20 rates are identical, about 7.5%, would require 50,000
21 patients in each treatment group. Up until then, I was in
22 complete agreement. I don't know if one proposed way of
23 getting around this is just to assume that equivalence means
24 a bigger difference because one proposed approximation is

1 the use of a larger delta, 1.5%, to decide that the
2 innovative therapy has provided sufficient evidence of
3 efficacy when tested against an active control. You could
4 quite easily end up concluding that an ineffective treatment
5 was equivalent if you took this approach. But you can see
6 that really the numbers are big.

7 (Slide)

8 So, my three concluding slides -- here is the
9 COBALT result, and I just wanted to touch very briefly on
10 one additional problem. If we combine COBALT plus the FTT
11 comparison of fibrin therapy versus control, then our point
12 estimate for double-bolus tPA versus nil is 19%, but it is
13 pretty close to zero.

14 But you could combine it in different ways. You
15 could say, well, the fibrinolytic trial overview combined SK
16 and tPA and in the study design they were basing their power
17 on the comparison of bolus tPA versus accelerated tPA. We
18 need to put in the difference between SK and tPA. Well, you
19 could do that in different ways. You could do it by saying,
20 well, we have three large trials that have compared
21 streptokinase versus tPA; they should be combined. So, we
22 are now going to do a number of indirect comparisons: bolus
23 tPA versus accelerated tPA; tPA versus SK; fibrin therapy
24 versus controls. There are three sets of random errors.

1 Or, you could say, well, I don't accept this. I
2 think that the only appropriate comparison of streptokinase
3 versus accelerated tPA is from part of GUSTO-I, which is I
4 think a perfectly appropriate thing to conclude. If you
5 wish to do that, then you can come up with a different
6 estimate. So, you can conclude double-bolus tPA may do
7 nothing, or that it may produce something like a 9%
8 proportional reduction.

9 Again, the difficulty is in which things you
10 include and whether you want to have a number of indirect
11 steps with greater random error. You can see the standard
12 deviations are increasing as you put in extra steps.

13 (Slide)

14 So, my tips on concluding falsely that new and
15 standard treatments are equivalent -- first, overestimate
16 the differences in outcome between standard and no treatment
17 because then you will be able to conclude that a big
18 difference between the standard and new treatment is really
19 equivalent, and is better than nothing.

20 Ignore the impact of the many different sources of
21 statistical and clinical variation on the estimated effect
22 and the treatment.

23 Study patients in whom the standard treatment
24 produces small absolute effects and low risk individuals, if

1 you are looking at absolute differences, or proportional
2 effects. If you want to make treatments look more similar,
3 then take patients where the standard treatment doesn't look
4 to be that effective.

5 Assess the differences in surrogate outcome
6 measures because the problem is that although a surrogate
7 may be associated with outcome, changes in the surrogate
8 with treatment may not be associated with changes in the
9 outcome.

10 Compare composite outcome measures that include
11 events influenced by the standard treatment only marginally
12 -- my example of adding on sort of recurrent angina in
13 fibrinolytic trials, or those that are affected in different
14 directions.

15 Then I think the best one is do positive control
16 superiority trials to detect unrealistically large effects.
17 Then a lack of significant difference implies equivalence.

18 (Slide)

19 The final slide -- sorry to have run on a bit
20 long. So, I think the key message is that we are not trying
21 to demonstrate equivalence; we are trying to demonstrate
22 that the treatments are effective and they retain sufficient
23 effect that they are worth having. Add-on studies are a
24 much better design. They will likely be smaller than an

1 equivalence study because the differences will be bigger.

2 Positive control studies need to be much larger
3 than null control trials. They are not a substitute, as I
4 stated, for mega trials. They are actually a requirement
5 for mega trials. The combination of proportional estimates
6 from positive control, null control studies may be more
7 generalizable but they may represent equivalence in some
8 settings and not in others because proportional effect in
9 high risk individuals will, in absolute terms, be bigger
10 than proportional effect in low risk individuals.

11 Separate estimates of effects on particular
12 adverse outcomes may be more reliable, and they may be more
13 generalizable than assessment of composite outcomes to
14 people with different risks or different proportions of
15 their adverse outcome due to stroke or death in that
16 particular circumstances, and surrogates will certainly not
17 suffice.

18 Thank you.

19 DR. PACKER: As the Committee is reorienting
20 itself, Rory, you mentioned an important misconception which
21 is that some of the enthusiasm for equivalence trials is
22 based on the fact that they are smaller and, therefore, more
23 doable. You have made the point that, in fact, a true
24 examination of the question of equivalence, in fact,

1 requires a substantially larger trial in order to rule out a
2 significant difference comparable to the existing data base
3 for the active control.

4 DR. COLLINS: A larger trial and even perhaps
5 greater uncertainty because of the sort of historical nature
6 and now knowing whether equivalence means that in that
7 particular circumstance both treatments are ineffective, or
8 relatively ineffective.

9 DR. PACKER: What I would like to do is have some
10 discussion of this issue because the prevailing wisdom, or
11 lack of it, is counter to that conclusion, and I just wanted
12 the Committee to explore that more fully because if, in
13 fact, the conventional wisdom is incorrect that would be an
14 important message to send home from this Committee meeting.
15 So, is there any discussion about that? Rob?

16 DR. CALIFF: I couldn't agree more with most of
17 the points that Rory made. I think that this has been well
18 described. It is written up in a bunch of places. It is
19 very hard, I think, for clinicians to accept the reality but
20 the clear issue is how much uncertainty in the negative
21 direction we are willing to tolerate and still prescribe or
22 advocate that the treatment should be made available to the
23 public. I think the only reasonable conclusion one can come
24 to is that we do need much larger trials than we are used

1 to, and I will sort of leave it at that and make one more
2 comment.

3 I think it is very feasible to do much larger
4 trials, but right now the regulatory environment is such
5 that millions of dollars are spent on collecting useless
6 pieces of data and doing so-called regulatory things which
7 don't really contribute to the question that needs to be
8 answered for life-threatening diseases. For example, in
9 thrombolytic trials if one could do almost no monitoring of
10 the data and simply record whether patients were dead or
11 alive or had a stroke, and put the millions of dollars that
12 go into monitoring and flying people around to make sure
13 doctors are telling the truth into enrolling more patients,
14 you would get the answers that we really need. I think it
15 is really a tragedy the way things are being done.

16 DR. PACKER: Let me have the Committee focus on
17 the first half of what you said, Rob. The question, I
18 think, is to Bob Temple or to Ray or Bob Fenichel. From the
19 present regulatory perspective, there has been discussion in
20 other meetings that equivalence trials, to be persuasive,
21 can be smaller, and Rory has said, no, they need to be
22 bigger. Which view do you share at the present time because
23 the answer could be different depending on how the question
24 is phrased?

1 DR. FENICHEL: Well, we certainly see equivalence
2 trials all the time for bioequivalence, and the size of
3 those trials is typically 6 patients or 12 patients. And we
4 have a very simple definition of bioequivalence. That is
5 fine for the effect of getting drug into the blood stream
6 and measuring that effect. Most clinical effects are much
7 more difficult to determine and the calculations are really,
8 you know, statistics 101. I don't know where this ignorance
9 has come from. It is widespread, as you pointed out, but it
10 has not been supported by the Agency, except in this obscure
11 area relative to clinical considerations of bioequivalence.
12 So, no, it did not come from us.

13 DR. LIPICKY: I can only add to that that there
14 are positive control trials that are for some purpose that
15 when we see them, we tell the sponsors we don't care if they
16 do them or not. They are wasting their time doing them
17 because they don't address the issues of a positive control
18 trial that we have been talking about so far this morning.
19 And they are poorly conceived and they are small and they,
20 in fact, contribute very little except for safety
21 information because at least there is some control.

22 DR. PACKER: Bob Temple, do you have any comment
23 on this? Oh, he is not there; I guess I couldn't see.
24 Udho?

1 DR. THADANI: Rory, you put it very nicely that we
2 really need larger sample sizes for the equivalence trials.
3 This is also true for the possible adverse effects. And I
4 think that would be very important because not only are we
5 trying to define the treatment as equal but to protect
6 patients from possible adverse effects. And I think some
7 drugs are withdrawn when the sample size was not large and
8 the adverse effects were not detected.

9 The other difficulty I sometimes have in looking
10 at the trial results is your issue, you know, of composite
11 endpoints. I think death and stroke is fine but then you
12 add on another, myocardial infarction or Q wave, and then we
13 don't know how much we miss where we do enzymes one time and
14 another time just plain no, and that creates more
15 uncertainty. So if you take that issue, would your sample
16 size have to go from 50,000 to 100,000, or what is your
17 opinion on that? Is that a major concern so that one should
18 say, alright, we can't make those primary endpoints; let's
19 look at secondary points and stay away from that, and just
20 stick with what we really truly can measure? I would like
21 your comment on that. For interpretation of results, I get
22 more and more confused.

23 DR. COLLINS: Well, I think the composite outcome
24 one is another little fraud that is going on. I mean, the

1 idea is to increase the number of events but it is actually
2 not increasing the number of informative events. If you are
3 adding in events that aren't influenced by the standard
4 treatment anyway, then it will make things look equivalent,
5 like recurrent angina as an example.

6 But even for sensible composite outcomes, if you
7 want to know about the balance of effect on death and
8 stroke, if they go in opposite directions then actually it
9 would be better. You would need smaller numbers to
10 determine equivalence by looking at them separately, than to
11 really be really assured when you combine them together and
12 then compare the two combined numbers. That was the example
13 I was trying to put up.

14 But I think this thing on equivalence trials and
15 people coming up and saying equivalence trials can be
16 smaller is based on doing fake power calculations where you
17 say I am doing an equivalence trial but I don't believe they
18 are equivalent. Therefore, to demonstrate equivalence, or
19 at least as effective, I only need X thousand patients. So
20 you now have wide confidence intervals on what looks like a
21 reasonable power calculation. So, with a bit of luck the
22 point estimate now is all in the luck of the gods; it is all
23 a play of chance. If it is a little bit better in the new
24 trial, then you can say, well, it is at least as effective.

1 If it is a little bit worse, you say, well, it is
2 equivalent. It is actually very difficult to make it look
3 worse. It is brilliant.

4 DR. PACKER: Rory, the other complexity one gets
5 into in a composite endpoint is not that the components of
6 the composite go into opposite directions, but they could go
7 in the same direction but be influenced in two different
8 magnitudes.

9 DR. COLLINS: Yes.

10 DR. PACKER: To the extent that one includes a
11 component which has a weak treatment effect or zero
12 treatment effect -- it doesn't have to be an opposite
13 treatment effect -- one is enhancing the ability to show
14 equivalence.

15 DR. COLLINS: Yes, like recurrent angina, which
16 would perhaps have no effect. I mean, there is obviously a
17 spectrum in between.

18 DR. SEIGEL: I would like to address this issue of
19 the confusion about whether equivalence trials are smaller
20 or larger. A lot of the confusion arises from the fact that
21 people don't specify compared to what.

22 Regardless of where you set your estimate of drug
23 effect, showing something by comparison to an active control
24 is going to require more patients than comparing to placebo,

1 and leave less certainty, at least determining drug effect,
2 will require more patients always in the equivalence trial,
3 and the equivalence trial will have to be larger than the
4 placebo-control trial.

5 But if your intent is to consider only a
6 comparison of your drug with standard therapy, and you
7 believe your drug to be superior, and your choice is either
8 to set out to demonstrate that it is superior or, rather, to
9 set out instead the less stringent thing that, at worse, it
10 is not much inferior to, then that is a less stringent thing
11 to do. That comparison, that same comparison takes fewer
12 patients. And if you can get on the market with that
13 comparison, and if you are going to do an active control
14 trial anyhow then, in fact, the equivalence active control
15 trial is smaller than the superiority active control trial.
16 But it is always larger than the placebo-control trial.

17 DR. COLLINS: I would like to comment on that
18 because you used the words "if you believe that your
19 treatment is superior." What is the basis for your belief?
20 I mean, is it appropriate that your belief should influence
21 everybody else's belief?

22 DR. SEIGEL: Well, obviously, what you believe is
23 always the basis for the size of the trial you do, and the
24 nature of the trial you do.

1 DR. COLLINS: But if you are testing equivalence,
2 surely your belief should be that they are equivalent and
3 your power then should be determined based on that, not
4 based on your belief that it is superior.

5 DR. SEIGEL: No, I don't disagree with that at
6 all. I am simply saying if you have a new and better drug
7 and you set out to do a superiority trial, someone will come
8 along and say, well, you can actually do an equivalence
9 trial for less. That, I think, is the source of confusion
10 in saying that equivalence trials are smaller. It is only
11 that limited application.

12 DR. TEMPLE: Whether it is better or not all comes
13 out in the wash. If you are, in fact, better a smaller
14 study will be able to exclude the margin that you said you
15 have to exclude to declare equivalence. So, it doesn't
16 matter. You are not foisting that on the rest of the world;
17 you are just choosing a sample size. If you are wrong, you
18 will fail to show equivalence and then you lose.

19 I don't know if this has come up, but an
20 equivalence trial of any kind is always going to be bigger
21 than a placebo-control trial because, at a minimum, you have
22 to choose a very conservative estimate of the control
23 placebo difference, whereas, in a placebo-control trial you
24 have to take the most conservative possible assumption about

1 how big the control placebo difference is and then show that
2 you are excluding that, and it is always going to be easier
3 to beat a placebo. So, I guess I don't know where the idea
4 comes from. It can never be smaller than a placebo-
5 controlled trial.

6 DR. D'AGOSTINO: You know, I can't speak for the
7 wisdom of the cardiovascular-renal community. In the
8 broader arena when people think about setting up trials, in
9 my sort of view, somewhat from the bioequivalency notion of
10 how easy those are in terms of sample size but also from
11 setting up the idea of the null hypothesis being equivalence
12 and the alternative being superiority, and it is somewhat
13 equivalent to what Rory is saying, that you accept the null
14 hypothesis and then you say, well gee, the drugs are
15 equivalent.

16 But if you actually do an equivalency trial things
17 are reversed, and a lot of people in the field aren't really
18 aware that things are reversed. Maybe I can show something
19 on the board here?

20 DR. PACKER: Go ahead.

21 (Slide)

22 DR. D'AGOSTINO: In the sort of standard theory of
23 hypothesis testing, you basically set up a couple of
24 treatments, say, that are equal against that they are

1 different, and you look at your sort of hypothesis test
2 being basically that as long as your statistic came out on
3 one of the extremes, you reject the null hypothesis and if
4 your statistic, be it either odds ratios or main differences
5 or what-have-you, came in somewhere in between you accept
6 it. A lot of people that I deal with think that when you
7 accept you are talking about a equivalence. The statistics,
8 when they talk about equivalence, are really setting up
9 something where the first drug exceeds the second one at
10 some ratio and you want to do two tests of hypotheses.
11 Basically, one is that you want to show that under your null
12 hypothesis, under your first null hypothesis you are in this
13 area versus the alternative, in this area. So, basically
14 one drug isn't better than another by a delta. Then you
15 want to do a second hypothesis in the opposite area saying
16 that the other drug isn't better by a delta. You basically
17 have to end up rejecting two hypotheses in order to make the
18 equivalence.

19 That feature, I am afraid to say, has not caught
20 on with a number of people. They are thinking this and
21 these samples sizes could be quite easy; not thinking of
22 this where the real equivalency trials are actually
23 substantially larger. I think a lot of the vocabulary
24 hasn't caught up again. Again, I can't speak for the

1 cardiovascular community but I can speak for all other
2 people that I deal with, and these two notions aren't really
3 clear.

4 DR. PACKER: Can I ask one follow-up question to
5 what Rory and many others have said and what Ralph is now
6 emphasizing? Is it possible for a sponsor to propose a
7 trial that is described as an equivalence trial in which the
8 intent is not to show equivalence in the way that Ralph has
9 now defined equivalence, but to show that the drug is
10 actually better than the putative placebo because there is
11 reasonably good data on what the comparer will do? I hope I
12 have define that clearly enough. In other words, it follows
13 from Dr. Seigel's comment. One could actually propose a
14 standard which is substandard to equivalence, but which
15 would reasonably be equivalent to beating placebo.

16 DR. COLLINS: That was really the message I was
17 trying to get across. One is actually not interested in
18 equivalence per se. One is interested in determining if the
19 new treatment is more effective than not giving it. You
20 can't do it directly unless you do an add-on study, and if
21 you want to reduce the sample size that is the way to go.
22 You can't do it directly; you have to do it indirectly. But
23 you could quite appropriately conclude that a new treatment
24 was not equivalent to a standard treatment but it was better

1 than no treatment.

2 If it was much cheaper or much more convenient, I
3 would say approve it. I don't think that things have to be
4 equivalent to the standard treatment. There may be other
5 advantages. What you want to know is that they are
6 effective to a worthwhile extent, and you are having to do
7 it indirectly.

8 DR. PACKER: Rob, would that bother you a lot if
9 someone did that?

10 DR. CALIFF: The key words were at the end there
11 from Rory. I have forgotten exactly what he said but it
12 wasn't just better than placebo but worthwhile. I think the
13 definition of worthwhile involves judgment about how the new
14 treatment stacks up against the conventional care. In other
15 words, if it was incredibly cheaper -- you know, if the
16 standard treatment is \$2000 and the new one is \$100, in
17 today's society we can't pay for everything so you would
18 expect a fair amount of loss of life, for example, under
19 that situation potentially.

20 DR. COLLINS: Take an example that may be real.
21 Let's say that all the trials that have been done for fibrin
22 therapy versus control were tPA, and someone came along with
23 this drug, streptokinase, and they wanted to get it
24 approved. And to take away argument, you had two arms of

1 GUSTO where there is a difference of about 1% between SK and
2 accelerated tPA. If you combined that with all of your
3 previous tPA trials, you may well conclude that
4 streptokinase is effective. You would conclude perhaps that
5 it is not as effective but you would conclude that it is
6 effective and you might approve it.

7 DR. PACKER: But in the United States we would
8 approve it even if it was twice as expensive instead of 5%
9 of the cost.

10 DR. COLLINS: Yes, but I am taking a real example
11 of real data, and I think it would be an appropriate thing
12 to have on the market and people are using it.

13 DR. KONSTAM: I really like what you are proposing
14 personally, and I would urge, you know, heading in that
15 direction. That is, if the key is saying this drug is
16 effective, then when you are designing your active control
17 trial what you really want to do is design it so that it is
18 different from the putative placebo.

19 Now, I think once you say that you get into the
20 next problem, and this has really permeated the whole
21 discussion right from the beginning in the background. The
22 difference between community standards and ethical
23 constraints that are perceived in the community and
24 regulatory perspective about what is done in the community

1 and whether it really works. I see this as an enormous
2 problem because we see this in some of the things we are
3 going to be considering over the next day. We have seen it
4 with enoxaparin versus heparin.

5 I guess there are two ways to go about approaching
6 it. One is to say let's just forget about it. You know,
7 our regulatory standard is our regulatory standard and at
8 the end of the day we are going to have to decide, on that
9 basis, whether the active control is or is not efficacious
10 independent of community judgment, whatever that is. Or,
11 you can say, you know what, we have a big problem out there.
12 There is a lot of consideration out there that placebo-
13 controlled trials in certain circumstances are unethical
14 despite the fact that the active control has not received
15 regulatory approval. I, for one, would like to urge the FDA
16 to really deal with this problem and proactively say what do
17 we do when there is community practice that is widespread
18 that has not rigorously reached our regulatory standard.

19 DR. LIPICKY: Like carrying dopamine around on
20 your back? I don't understand what you said.

21 DR. KONSTAM: I don't have an answer to it, Ray.
22 From a strict perspective, I am very much in favor of
23 adhering to strict regulatory guidelines. I want to do
24 that. The question that I see is that this is going to come

1 up again and again, and it came from one of the very first
2 slides of Bob's that you challenged, which is the difference
3 between community standard and regulatory standard. I mean,
4 I don't know what to do. I don't have an answer but I guess
5 if we say forget it, you know, if it hasn't passed our
6 strict regulatory perspective there is nothing we can do
7 about it.

8 DR. LIPICKY: See, Bob clarified that pretty well
9 in the sense that it isn't what FDA has approved. The issue
10 is not regulatory standard but how one will make the
11 judgment. If, in fact, you have a treatment that has never
12 had a treatment effect demonstrated, how can you evaluate a
13 positive control trial? You know, that is not a regulatory
14 standard. That is not the issue. And if you want to say
15 how you can tell that magnitude of treatment effect that you
16 want to preserve, never knowing what that treatment effect
17 has been, we are willing to listen.

18 DR. PACKER: Marv, I think there are three
19 standards. One is a regulatory standard. We are familiar
20 with that. The second is the community standard. I think
21 that no one on this Committee is suggesting that an active
22 control against what the community thinks is acceptable is
23 remotely acceptable. There is a third category, which is
24 that there are some drugs for which there may be persuasive

1 data for which no sponsor has actually filed an application,
2 and, yet, the data are truly persuasive. I think what Bob
3 Fenichel was saying is that that is the ideal active control
4 because even what the FDA approves may not necessarily be a
5 sufficient criterion if it doesn't meet in itself a
6 consistent persuasive standard. So, it is the presence of a
7 consistent persuasive standard which overrides all
8 categories. I think I am summarizing that accurately.
9 Right?

10 DR. FENICHEL: yes, that is what I said and that
11 is also what Dr. Temple said.

12 DR. TEMPLE: Well, for one thing, it is going to
13 be relatively rare for a treatment to be good enough for you
14 to say that it regularly beats placebo and not be in any
15 labeling anywhere. That happens but it is not going to be
16 very common.

17 It sounds to me like there a couple of things
18 ought to be teased out. One is an ethical concern, and
19 another, and completely separate, is whether an active
20 control trial is interpretable.

21 I would assert, for example, that you can't do
22 infarction beta-blocker trials with a new beta-blocker any
23 more because through meta-analyses and individual studies we
24 know that beta-blockade is life-preserving after heart

1 attack. I would also assert, however, that an active
2 control equivalence trial is uninformative because most of
3 the trials of beta-blockers have shown a benefit. Now, that
4 could be because they have been too small but until somebody
5 does the large trials, which no one ever will, I can't ever
6 know that. So, the community would say, and I would agree,
7 it is an unethical trial. I would also say that an active
8 control equivalence trial cannot be informative so you are
9 stuck.

10 That raises a point that Rory addressed, can you
11 do an add-on trial? Well, if you want to show another beta-
12 blocker is effective like timolol or propranolol, I would
13 say there isn't any add-on trial that is informative about
14 that. If you want to find out whether some new
15 pharmacologic intervention can give you even better survival
16 after a heart attack, of course, you can do an add-on trial,
17 and we spend all of our time urging add-on trials in
18 oncology, for example, where it is always more interesting
19 to see if you can do better than if you can do just as well.

20 The other thing that came up is suppose you meet
21 the standard for equivalence by showing that the difference
22 between you and the control isn't larger than a certain
23 amount but you are actually inferior. That is theoretically
24 possible but practically extremely unlikely. You design

1 your trial to be big enough so that if you are equivalent
2 you will be able to exclude the difference of interest. To
3 squeeze into a trial like that the possibility that you are
4 actually demonstrably inferior and still superior is very
5 difficult in all but the most unusual circumstances.

6 Where the difference between no treatment and
7 treatment is very large, like in antibiotics, you can
8 actually do that. You can sometimes show that one
9 antibiotic is inferior to another and, yet, you are quite
10 sure it actually has an effect. That raises something of
11 the problem you described. It is not easy to see how you
12 could do that though in a large trial where you are
13 straining for numbers in the first place. That doesn't mean
14 the point estimate couldn't be slightly below but that is
15 hardly the same as inferiority. That is just a point
16 estimate that is slightly low.

17 But to actually, you know, run a trial that is so
18 big that you could show that streptokinase is better than
19 placebo but is inferior, I think it would be very unusual to
20 be able to do that. It is not that it couldn't happen; it
21 is just that the numbers would have to be so vast.

22 DR. PACKER: Bob, I know that many members of the
23 Committee would like to comment, and I think we have the
24 general issue for discussion as to the question of whether

1 beating placebo is a regulatory standard, or providing
2 reassurance of similarity to an existing therapy -- those
3 are very important issues, and what I would like to do now
4 is to thank Dr. Collins very much. We are going to take a
5 twenty-minute break and then begin again with Dr. DeMets'
6 presentation straightaway after the break.

7 (Brief recess)

8 DR. PACKER: There are a number of important
9 issues that have been brought up this morning, and we will
10 try to explore as many as we can as the morning proceeds.
11 let me again emphasize that the purpose of this morning is
12 not to reach specific decisions but really to provide an
13 opportunity to explore the issues, and to get a sense
14 perhaps more of what we should not be doing than perhaps
15 what we should be doing, although hopefully we will get some
16 insight on the latter as well.

17 So we will proceed with Dave DeMets. Dave, thank
18 you very much for being here. I don't know who made up the
19 titles but your title is, "If these are the Circumstances,
20 This is How to Calculate Things." It sounds like a Broadway
21 show.

22 **If These are the Circumstances, This is How**
23 **to Calculate Things**

24 DR. DEMETS: It is a fascinating title and I am

1 not sure I am going to deliver that one. At any rate, I
2 want to talk about some of the quantitative aspects of the
3 problem. Many of the issues have been alluded to already
4 during the course of this morning's discussion. So, in some
5 sense, as with any speaker down the list, things have been
6 discussed that you intended to say but I will say them
7 anyway briefly.

8 (Slide)

9 Actually, I borrowed this figure from a paper that
10 Tom Fleming wrote for an AIDS conference that we were at
11 several years ago. But it does get to the issue that Dr.
12 Collins was mentioning about add-on, or what I call
13 classical where you add an experimental new therapy to a
14 standard. In the active control, you compare the active to
15 the standard and you could, in fact, try to show superiority
16 in that design, or you could try to show what is called
17 equivalence. I think we just need to keep those factors in
18 mind as we go on.

19 (Slide)

20 I think it has already been implied, but it is
21 certainly true that superiority trials are difficult and
22 challenging enough but the equivalence trials are even more
23 challenging. At least, I will try to raise a couple of
24 issues that I don't think have been described so far this

1 morning.

2 (Slide)

3 In any trial the noise factor is what you are
4 trying to beat, and in a superiority trial you, obviously,
5 have a strong incentive to minimize the noise because you
6 are trying to detect something. If you are not careful in
7 an equivalence trial, the noise factor is, in fact, going to
8 work in your favor. I will talk about that but, you know,
9 adding patients that are ineligible, losing data or losing
10 track of patients, noncompliance, just general sloppiness,
11 and I am at least going to talk about the noncompliance
12 implications a little bit later.

13 (Slide)

14 What I thought I would do is show some of the
15 parallelism and contrast between classical superiority
16 trials and the equivalence trials, and hope that I don't
17 insult anybody here by taking this simple-minded approach.
18 But I can get lost in some of the language that has been
19 used so I will try to go through it simply.

20 In the classical experimental situation we talk
21 about the null hypothesis, no difference in response in the
22 two groups. That means that the delta in response is zero.
23 And we specify some alternatives that we expect to see, hope
24 to see and would like to see. In order to have some kind of

1 design parameters we talk about a significance level. By
2 that, we mean a type I error or false-positive rate that is
3 claiming that there are differences when, in fact, there
4 aren't any. That would be an error that we would like to
5 minimize.

6 The second design parameter has to do with the
7 other kind of error, failure to claim differences when there
8 are, or sometimes we talk about power that is the
9 probability of rejecting a hypothesis, given that the null
10 hypothesis is not true.

11 One of the issues that was brought up this morning
12 in relation to power, power is a functionally specific
13 alternative. When you say you have a powerful study, that
14 might be true. It might be quite powerful but it is
15 powerful against an alternative that is humongous,
16 unrealistic. So, you can say I have a powerful study but
17 power is a function of the things that you have specified in
18 your design. So, we have to keep in mind in an equivalence
19 or non-superiority design that there is some difference that
20 we are thinking about, and it is that difference that we
21 have to have power for, otherwise the other power largely is
22 not useful.

23 (Slide)

24 So, we talk about type I error in a superiority

1 situation of 5%, 1%, maybe we talk about even more extreme
2 than that, and that the power should be at least 80%. I
3 don't think any of us would be interested in something that
4 was less, although many people still do trials that way but
5 I doubt we would invest our own money in that kind of study.
6 And we specify some delta that is at least the one we hope
7 to see. If we are doing this in the right way, it is the
8 smallest delta we hope to detect that is clinically
9 relevant.

10 The issue here is that if you are doing a
11 superiority trial and you have low power even for the delta
12 that you are after, the worst that happens is that you
13 missed finding something. But in a superiority trial if you
14 don't have power against something you can actually claim
15 something, that is that effectively one trial is as good as
16 the other. I will come back to that, but the issue that we
17 have to think about is whether the significance level and
18 the power in a superiority trial will show a similar thing
19 in the equivalence trial.

20 (Slide)

21 So, I want to get a little specific here for a
22 minute to illustrate some issues. If we are thinking about
23 a superiority trial where we have to event rates, failure
24 rates let's say, and we are going to compare those two rates

1 by a standard normal sample size that is large enough to
2 justify that, and we make the usual assumptions that the
3 sample sizes will be randomized equally, although that
4 doesn't change the argument at all, and we specify some
5 alternative that we are after, a difference hoped for, a
6 minimal clinical difference we hope for, then when we come
7 up with a sample size formula, which you have all seen, that
8 looks like the following.

9 (Slide)

10 The issue of type I error is represented in this
11 coefficient. You have what is essentially a variance term
12 here and a difference term, here. Now, it has been implied
13 all morning that one of the problems with showing
14 equivalence is that you can't show that the delta is equal
15 to zero for obvious reasons. That is an extremely large
16 trial. Not even the DUCS group, I believe can do that
17 trial, or Dr. Collins in the U.K. But this delta is a
18 critical issue in specifying power -- what delta you are
19 after, with what kind of power. Variability has also been
20 alluded to. I think we need to minimize the variability.
21 The proportion in variability is a function of the event
22 rates partly. It is also a function of patients but,
23 strictly speaking, variance is a function of the event rate.

24 (Slide)

1 If you specify some difference -- I have plotted
2 here the total sample size in a trial versus the function of
3 the event rates, and this is on a scale of reduction of the
4 ratio of the two event rates. If you had a 25% reduction
5 that you are looking for in a superiority trial you might
6 end up with a sample size slightly under 1000 patients, with
7 a two-sided alpha of 0.05 and 90% power.

8 Relative to this morning's discussion, the active
9 control you pick or, in this case, the placebo event rate
10 you have has a lot to do with the sample size. If you are
11 picking an active control, which active control you pick
12 matters because it will have something to do with the event
13 rate. That will imply a larger or a smaller sample size.

14 Second of all, my experience is that event rates
15 change on you from one trial to the next. Even when you
16 think you have the same population, the exact same treatment
17 and perhaps the exact same dose but you are doing it again
18 later for some reason, surprisingly event rates change on
19 you. We will come back and talk about that a little bit
20 later.

21 (Slide)

22 The issue of noise -- one source of noise in a
23 study is the issue of noncompliance. Non-compliance can be
24 manifested in several ways. But if you take the intent-to-

1 treat principle which Dr. Collins talked about, a simple --
2 you can get a lot fancier than this, but a simple estimate
3 of how much impact noncompliance can have on your study is
4 by looking at the noncompliance rate and adjusting the
5 sample size if you had perfect compliance. You adjust this
6 factor and, so, if you had 5% noncompliance, in order to
7 keep the same power, you have to increase the sample size by
8 10%. If you have 10%, you have to increase the sample size
9 by 0.3% and so forth. If you have kept the sample size the
10 same and didn't change it, the power is going to drop off,
11 something like 85% and maybe 80%, and less than 50% or 60%,
12 down here. So, the noncompliance is going to have a big
13 impact on the power that you really have. If you don't
14 account for that in the design you will have an under-
15 powered study even if you think it is pretty powerful. That
16 is true for superiority trials and it is certainly true, and
17 probably even more critical for non-superiority trials.

18 (Slide)

19 Probably the best way that we like to look at
20 results is through confidence intervals, and most authors
21 who write about non-superiority trials or equivalence trials
22 think of it in terms of the confidence interval approach.
23 Here you can also see, whether you are looking at the event
24 rate itself and the standard error as a function of

1 variability and the sample size, or you are looking at
2 relative risk, how many standard error difference do you
3 think is important and meaningful and you have the
4 variability. The issue of variability is going to be a
5 major factor in interpreting your data with a positive
6 control trial or a classical control trial.

7 (Slide)

8 What we do with confidence intervals if we have a
9 superiority trial -- this is sort of experiment one, we have
10 the placebo rate and some function around that, the
11 mortality rate and if these confidence intervals overlap we
12 would say that they are not significant by the standards
13 that we have set for ourselves. One can be a little more
14 efficient, I suppose, by looking at the differences but for
15 today's purposes to demonstrate the issue, I have kept the
16 two rates separate and not looked at the difference.

17 If you have this situation, in experiment two, you
18 have the active treatment clearly and the confidence
19 intervals don't overlap and you would claim that there is a
20 difference.

21 I also looked at the relative risk. In experiment
22 one, if you include one; experiment two, exclude one. So,
23 you would claim a difference or you wouldn't.

24 I think using a confidence interval gives you a

1 lot of information about the experiment that you have, and
2 you can tell where the estimate is and how much you know
3 about it. If this is a very tight estimate you feel better
4 about it; if it is very wide you feel less well about it.

5 So, superiority trials -- is this sort of the
6 paradigm we have all sort of worked in a lot? As has been
7 said, often there are two goals. Sometimes you try to
8 accomplish them in the same study. Often you want to show
9 equivalence by saying that the experimental treatment is no
10 worse than the active by a certain amount. Sometimes we set
11 up for superiority but if it didn't make that, well, you can
12 certainly go for an equivalence trial. We might want to
13 talk about whether that is a good idea or not.

14 (Slide)

15 So, maybe you are seeking equivalence, maybe also
16 superiority and, given my own history, just my recent
17 history, I always think that there is a possibility of harm
18 and when you are pitting two experiments against each other,
19 goodness knows which way they are going and I think you have
20 to at least think about the fact that things might go in the
21 wrong direction by a given amount.

22 Others have said several times, and I think there
23 is a lot of confusion, misunderstanding, that to reject a
24 null hypothesis does not constitute equivalence, and 80%

1 power isn't adequate and noncompliance is a serious issue.

2 (Slide)

3 The way most authors who write about this, which
4 is what Dr. D'Agustino was getting to, is that the paradigm
5 is sort of flipped and in a superiority trial you are trying
6 to show that there is a difference; the null hypothesis is
7 that there is no difference. Although one, in fact, usually
8 uses the criteria of zero but, in fact, there is nothing
9 that says you couldn't specify some small difference. You
10 have to beat not just zero but some small amount. We tend
11 not to do that but we could.

12 But in superiority trials you reverse those. The
13 null hypothesis is that there is a difference more than some
14 delta, and you are trying to show alternatively that it is
15 less than that. So the classic references that sort of
16 talked about this early on were Bill Blackwelder and Bob
17 Makuch and Rick Simon. So, this concept has been around for
18 a while in terms of hypothesis testing and role reversal.

19 (Slide)

20 However, if you sort of follow through those
21 details in terms of design principles, you wind up
22 essentially the same or very close to the same sample size
23 considerations. The roles of these two coefficients get
24 turned around but they are both there and you have to decide

1 how much error of each kind you want to make, but clearly,
2 for me, the bottom line is to keep it simple and not twist
3 it around in language, if you want to have a lot of standard
4 errors or criteria which you believe is real, a large
5 probability of finding the difference is the delta. So, the
6 focus in what difference are you looking seems to be where
7 most of the decisions are going to fall.

8 (Slide)

9 So now the issue which has been raised is which
10 active control. You may have a couple of choices here to
11 make. An active control which has a big effect, is one set
12 of implications in terms of the event rate because you are
13 now going to go against one of these two. And the most
14 effective one has a smaller error rate in this case. As you
15 remember from that earlier slide, the lower the event rate,
16 the tougher the job. So, it does matter which active
17 control you pick. Even if it is well-established for being
18 better than placebo, it can have an effect on your designs.

19 (Slide)

20 Now, I had another figure which I have modified
21 from this paper which Tom Fleming did on AIDS a few years
22 ago, contrasting the two situations. I have it on a scale
23 of relative risk. In this sense, the relative risk is
24 bigger than one if it is harmful; less than one is

1 beneficial. Placebo is at one. You specify some delta you
2 think that the treatment can be improved, you specify that
3 delta, and you look at what you have got and if the
4 confidence interval is larger than one, if the lower limit
5 is greater than one you would claim harm. If you are
6 somewhere in between you would say it is not significant.
7 You wouldn't claim equivalence. And the issue that has been
8 already raised, if you are less than one in the upper limit
9 you would claim benefit.

10 In the active control what happens is that now you
11 are shifting and the standard is not the active on the
12 relative risk and it is a standard against itself. And, if
13 you are using your new therapy and it turns out that the
14 confidence intervals were greater than one, you might claim
15 it is worse. If you are thinking about what delta you want
16 to specify that it is no worse than, maybe you use the
17 standard estimate and its confidence intervals, plus/minus
18 two standard errors, as your choice of delta. We already
19 heard discussion this morning that maybe that is not good
20 enough; you want to have it tighter, a tighter delta or
21 maybe a bigger delta. But whatever the delta is that is
22 chosen, it should be based on some sense, I would think, of
23 the estimate of this effect and the standard error. If you
24 can rule out this value by its upper level, the upper

1 confidence interval, you would say it is co-equivalent.
2 That is, it is no worse than the active control. Of course,
3 the best of all would be that the confidence interval upper
4 level excludes one.

5 This approach I think would give us a lot more
6 information than talking about the classical hypothesis
7 testing and trying to keep track of which direction our type
8 I and type II errors are in.

9 I want to come back to this picture in a minute
10 and also go on to the issue of placebo. We had a lot of
11 discussion this morning about this. Which active control
12 you pick or how many studies you pool together will position
13 the placebo event rate for the relative risk on this scale
14 relative to your standard. Maybe you want to draw
15 confidence intervals around that estimate as your criteria
16 for what you want to show, not to show how much am I worse
17 than the standard by a certain amount but how much am I
18 better than placebo. But it is very much an effect of which
19 active control you pick as to what the placebo rate would be
20 relative to that.

21 (Slide)

22 In addition, the issue of the active control --
23 people have sort of argued that you need to have both of
24 these issues, some estimate relative to the active control

1 and some estimate of placebo effect relative to the standard
2 you have chosen, either by a point estimate or a confidence
3 interval.

4 (Slide)

5 The problem with that, as I see it and as has
6 already been discussed this morning, is that that is a
7 moving target. I don't know how to get around it. The job
8 isn't necessarily to solve the problem but to raise the
9 issue. Why is it a moving target? Well, the disease
10 process is continuing, maybe not rapidly but it may be
11 changing in some sense. I think that is true in cancer and
12 I think it is true in cardiology. The background therapy is
13 changing. So, even if you were to do the exact experiment
14 all over again that you are basing everything on, it
15 wouldn't be the same experiment. If you took the same drug,
16 the same protocol and ran it again, the background rate is
17 likely to be different because of the background therapy
18 and, something that is very hard to quantify, but there is a
19 selection bias. It is very interesting how you think you
20 have everything the set and you just wrapped it up; you have
21 your event rate in the so-called control arm; you start the
22 protocol up and, lo and behold, you find out that the event
23 rate is less than you expected and that you just saw in the
24 last study because patients are selecting themselves, and so

1 is the healthcare system and the healthcare providers.

2 So, to me, it is very difficult to ask the
3 question with the active control against a placebo. It is
4 very difficult to figure out what the effect would be in a
5 new study. So, I find it very challenging and puzzling. I
6 don't have a solution to this but --

7 (Slide)

8 -- the problem I am focusing on here is that it is
9 difficult to figure out, first of all, which studies to put
10 in, as Dr. Collins pointed out. If you put them all in you
11 get a tighter interval. If you put in just the ones that
12 are the most relevant you get a wider interval. But even if
13 you did that relative to the trial you are doing today in
14 the context of today's healthcare with the patients who are
15 volunteering, it is very hard to figure out how relevant
16 that placebo event rate is. I know that doesn't solve the
17 problem but it makes it worse, but I think that is for the
18 judgment and the wisdom for the Committee. It is not a
19 statistical issue.

20 Thank you very much.

21 DR. PACKER: While the Committee is repositioning
22 itself, let me just ask you, Dave, you raised I think a new
23 issue which we have not dealt with yet this morning, which
24 is the issue of compliance. In the usual superiority trial

1 one does an intention-to-treat analysis, as Dr. Collins
2 mentioned, which is a conservative analysis, if one is
3 trying to raise the possibility of rejecting the null
4 hypothesis. But in a trial which will result in a claim of
5 equivalence an intention-to-treat analysis, if there is a
6 high degree of noncompliance, can be very confusing. The
7 confidence intervals that you are generating are based on
8 the number of events. Whether or not those events are in
9 compliant patients or in non-compliant patients, one could
10 conceivably have narrow confidence intervals which are
11 totally non-informative because the noncompliance rate was
12 very high. An extreme example would be if one carried out a
13 double-blind active control trial where there were 10,000
14 events in each group so that the confidence interval was
15 very narrow, but actually no one took the randomized
16 therapy.

17 How do you look at the issue of noncompliance
18 because it is a critical issue? We are usually comfortable
19 in being conservative in a study which is trying to show
20 superiority. But how can you possibly deal with this issue?
21 You can't deal with it by the narrowness of the confidence
22 intervals because those are event-rate driven. They don't
23 account for whether the patients have actually received
24 treatment.

1 So having raised the issue of noncompliance as a
2 horrendously complicating and confounding factor, can you
3 give us guidance as to how we deal with that, if not
4 quantitatively then, at least qualitatively?

5 DR. DEMETS: It is as important, if not more, in
6 an active control, so-called equivalence, trial to work
7 harder at the noncompliance issue than ever before, and to
8 make the trial as simple as you can. You can only go so far
9 with that. Obviously some patients won't comply totally.
10 It would be surprising if everybody did. But I think that
11 if you don't build into your design the fact that there will
12 be some noncompliance -- you have to have a certain
13 probability power to find that delta by whatever criteria
14 you have to be able to find that. If you don't adjust for
15 it your power is going to go down.

16 DR. PACKER: But since the power is based on the
17 event rate and the anticipated delta, both of which can go
18 according to plan perfectly well, it would be meaningless if
19 the noncompliance rate was very, very high.

20 DR. DEMETS: But power is the function of two
21 things. It is a function of the event rate, and it is also
22 a function of the difference. What noncompliance does is to
23 dilute the difference. It dilutes whatever difference is
24 really there.

1 DR. PACKER: It dilutes the difference, but if the
2 intent is to evaluate equivalence and the noncompliance rate
3 is very high, you are going to show equivalence even if the
4 therapies are non-equivalent.

5 DR. DEMETS: Right.

6 DR. CALIFF: Milton, I was going to amplify on
7 that. Based on what you said, if I was trying to take the
8 safest route as a sponsor to get on the market, I would pick
9 the worst of the active treatments already available and
10 give it in the form that you have to take it the most number
11 of times per day as my comparator. We are seeing that being
12 done. So you are giving an active drug that shows that
13 effect but you are maximizing noncompliance in the
14 comparator group and you are giving the least effective form
15 of the drug of the active control. Based on what you said,
16 that would maximize the chances of showing equivalence or
17 better.

18 DR. PACKER: You could that in lots of clever and
19 original ways, including having a drug that had a high
20 degree of side effects that requires withdrawal of the
21 study. I am just wondering, what is the conservative
22 approach to the analysis of the treatment effect in a highly
23 noncompliant patient population when the intent is to
24 evaluate equivalence? One knows the answer to that in a

1 superiority trial, but what is the conservative approach to
2 the analysis in an equivalence trial?

3 DR. DEMETS: Well, some of the creative things
4 that might be done, to me, have a high risk of introducing
5 bias. If you start looking the compliers -- let's just take
6 one example, I don't know what you are comparing. We have
7 plenty of examples to demonstrate that. There are all kinds
8 of biases for reasons we all know about. So the minute you
9 start tinkering around, taking people out, you destroy
10 randomization.

11 Other approaches people have tried to take have
12 been to do some modeling. But most of the models that I
13 have seen can break quite easily as soon as you say that
14 compliance is somehow a function of how a patient is doing;
15 it is not independent, which it probably isn't. So most of
16 the methods that people have tried I think are flawed. So,
17 you are stuck with the patient you have got. You can't get
18 rid of those. So the compliance is there. And the only way
19 I know to beat it is to minimize it.

20 DR. TEMPLE: Well, I am shocked by the cynicism
21 that I have heard that would suggest that people would
22 actually try to design trials that would show no difference.
23 You don't actually have to be that cynical. All you have to
24 do is notice, as Dave said and I think Bob said before, that

1 the incentives to producing a study that shows a difference
2 are lacking. Even forgetting about mortality trials, if you
3 think about a typical angina trial or hypertension trial
4 there is a period during which you make sure that people
5 actually have the disease. You exclude people who are too
6 variable because the measurements are no good. You have
7 lead-in periods to get rid of placebo responders. All those
8 things are designed to make sure you can show a difference
9 if there is one.

10 Why would anybody whose main goal is to show no
11 difference do any of those things? So in a million ways,
12 some which we are not even imaginative enough to think of,
13 the incentives to producing a different showing study are
14 missing. I guess I would put to Dave what do you do with
15 that? That goes to the location of your blue placebo dot,
16 and all of these things reduce the effect compared to
17 placebo. That is what they would all do.

18 DR. DEMETS: I think we have to attack the so-
19 called active control with the same vigor that we would a
20 superiority -- at the end of the day we want to be able to
21 say we are very sure we have done the best job and we
22 believe that we have ruled out that delta, whatever it is.

23 DR. TEMPLE: In symptom areas the Agency has
24 attacked it sufficiently that we have come under a fair

1 amount of criticism. I mean, we have attacked it so much
2 that you can't do it. There are no areas that involve
3 symptoms that I can think of where we accept equivalence as
4 being meaningful. I don't know if you have read the first
5 couple of paragraphs of Martha Angel's editorial on HIV
6 drugs but sort of casually and without paying much attention
7 to it, she asserted that if there is an existing therapy you
8 simply cannot do a placebo-control trial. Now, that was a
9 thoughtless comment and I am sure she probably wouldn't have
10 made it if she had thought about it more. But there was an
11 article in the New England Journal some years ago that said
12 exactly the same thing. So, people do say that sometimes.

13 But our position has basically been what you said
14 and in those areas where you can't be reasonably sure we say
15 they are not interpretable. What makes it difficult is
16 these areas where you are talking about mortality where you
17 can no longer do placebo-control trials. That is why this
18 discussion is so important. There becomes a major incentive
19 to try to figure out what you can do.

20 DR. DEMETS: I raised the issue a little bit about
21 if you have a trial with an active control. What do you do
22 if you take either outcome? If you got superiority you
23 would be delighted. If it failed you would take equivalence
24 if it ruled out some effect. In that trial you would have

1 the incentive to do as well as you could because being
2 superior would have an advantage.

3 DR. CALIFF: I was just going to comment, the one
4 thing, for sure, it seems like in the regulatory environment
5 you should do is not create rules which encourage people to
6 use the lesser effective active control in their trial. If
7 the goal is to beat a putative placebo, I think it is clear
8 to me that the current rules encourage the use of a less
9 effective active control.

10 DR. THADANI: On the compliance issue, I think
11 there are two issues. One is if you calculate your power
12 and the noncompliance is so bad, then you don't have a
13 trial. You could conclude that.

14 But the other issue is that compliance is poor
15 because a poor drug has some side effects and the patients
16 can't take the medication. You can't really force them
17 because they are having side effects. Then the question is,
18 is the data still valid because they are noncompliant
19 because of your adverse effects, not because they are not
20 taking the medication because they don't want. So what is
21 your impression on that, a noncompliance problem because the
22 drug could not be tolerated by patients? Say, you do a
23 study and he has a heart attack. You know, you might think
24 the study drug is producing it and I might try my best to

1 put him back on the drug but he is not going to take it. Is
2 the interpretation any different when you analyze the data?
3 Or, how do you tackle that? That is point one.

4 The other issue you raised is lack of adequate
5 follow up. What happens in some of the trials, once the
6 patient is not taking the medication it becomes a phone call
7 and then they lose interest and the follow up changes. And
8 the mortality trials are fine because you are counting
9 heads, but if there are infarct rates or other issues, you
10 could miss them if the patient doesn't come in. Could you
11 address those two issues and how to get around those?

12 DR. DEMETS: In the first one, where you say that
13 the standard or the active therapy has a lot of toxicity,
14 which would be typical in cancer, for example, where we have
15 a lot of toxicity, I think one thing you want to find out in
16 the trial I have just done is the noncompliance for toxicity
17 at least in the ball park of what is expected in all other
18 studies. If it is worse than expected, then I would worry.
19 If it is the same ball park, you would say, well, what I
20 want to do is get a trial that is maybe in the same
21 equivalence range, whatever that means, but the reduced
22 toxicity. Oncologists I think deal with this a lot. It is
23 probably not so prevalent in cardiology but certainly in
24 oncology. So, I think you have to find out if the

1 noncompliance rate that I am observing driven by toxicity
2 consistent with what we see in other studies.

3 DR. THADANI: One other issue comes up. People
4 say all right, because the drug may not be tolerable we are
5 going to look at the tolerability first and only include
6 patients who can tolerate the drug. But then you are
7 criticized because you are throwing out patients who
8 otherwise would have been in the study. Okay, you do two
9 weeks minimum tolerated dose which the trial is going to
10 involve but then you end up having some of the events during
11 that period of tolerability, and then the analysis becomes
12 complicated. So do you think those trials are good if you
13 look at the tolerability first before randomizing them, say,
14 to treatment A and B, or should that not be encouraged?

15 DR. DEMETS: I think what you are talking about is
16 having a run-in period --

17 DR. THADANI: For the active drug.

18 DR. DEMETS: Certainly, there are trials which
19 have done that. You know, it is valid to do it. You have
20 to understand what question you are asking. If you have a
21 run-in period and you exclude people who couldn't tolerate
22 the regime or the drug or the dose, you are asking a
23 slightly different question. You are asking does one
24 therapy beat the other in those patients who couldn't

1 tolerate the drug in a short period of time. It may be an
2 irrelevant question; maybe it is not. But you are asking a
3 different question. So if you agree with the question, it
4 is a valid way to approach it.

5 DR. THADANI: What about the follow-up period?

6 DR. DEMETS: The follow-up issue I think is
7 problematic in all trials, but it certainly is problematic
8 in active control trials because I believe that missing
9 data, or those kinds of issues, are independent of the event
10 process. I mean, they might be not perfectly correlated but
11 I don't believe they are independent. So, missing data goes
12 back to the issue that we should be very careful which data
13 we collect and just the right stuff. I do think we collect
14 more than we need, but the compliance to follow up is a
15 concern. You know, you don't want to be in a situation
16 where you are imputing data, imputing a placebo effect.

17 DR. LIPICKY: On the power business, there is a
18 reason to calculate power prospectively, that is, to decide
19 how big the trial is going to be and all that sort of
20 business. So, let's take the think that Milton outlined
21 when he first started off asking questions, this big, big
22 trial that has very many event rates and it, in fact, found
23 a difference between the two populations, and the difference
24 had a standard p value of 0.001. But, in fact, only 25% of

1 the treatment group took their medicines; 75% did not. So,
2 a retrospective power calculation would have said where the
3 power was originally something like 90%, 95%, it brought it
4 way down to 0.5 or something. Would that mean that one
5 should say the trial did not find something?

6 DR. DEMETS: No. Power after the fact can be
7 informative, but if you got a significant result you beat
8 the odds as you set them up. In the situation you outlined,
9 that therapy must be really fantastic because if you have
10 25% compliance and a p value of that size you really want to
11 examine this therapy very carefully. It did something even
12 in that noncompliant population.

13 DR. LIPICKY: Then the second question is Rory
14 said something, and I can't remember the name he associated
15 with it but it was like non-random error or something, when
16 you were saying you combined the results of three studies.
17 What was the name you associated with the error that gets
18 introduced?

19 DR. COLLINS: I was saying that you have to add in
20 the random error --

21 DR. LIPICKY: Random error.

22 DR. COLLINS: -- when you are combining, say, five
23 fibrinolytic versus nil and then one fibrinolytic versus
24 another and then one fibrinolytic versus another.

1 DR. LIPICKY: Well, was that thing you were
2 talking about in any of the equations that David showed?

3 DR. COLLINS: I think he was looking more at the
4 comparison within the positive control study. I feel that
5 the aim of these equivalence trials is actually not to
6 demonstrate equivalence but to demonstrate that the
7 treatment is effective and that one, therefore, has to also
8 include the statistical variance, as well as the uncertainty
9 or clinical variance. But you need to include the
10 statistical variance of your estimate of the standard versus
11 nil. And if you are doing it in a number of steps, standard
12 versus nil, newer versus standard, new versus new, then you
13 have a lot of variances.

14 DR. LIPICKY: But those are different from any
15 variances that were just talked about.

16 DR. COLLINS: Well, Dave was also talking about
17 variance around the placebo effect, which I suppose could be
18 considered in the same way as your trying to estimate the
19 variance around the standard, looking at it in a different
20 way.

21 DR. DEMETS: I just flipped it around. I mean,
22 Rory was talking about variance of the estimate of the
23 effect. I flipped things around where the standard is now
24 one and placebo is higher. But you can make that placebo

1 estimate as tight as you want, depending on how many trials
2 you dump in, and all the kind of variation you were talking
3 about is represented in there because of study variation,
4 what kind of patients, the size of the studies, and which
5 one is the right one is the tough question.

6 DR. MOYE: Just briefly to second what David said,
7 in the finding of a positive trial where you have very low p
8 value the issue of power really becomes meaningless. The
9 more delectable question I suppose is Ray's suggestion where
10 you have only 25% of patients in the active therapy taking
11 their meds and the p value winds up being 0.1. Then what
12 happens, because of course you have really an under-
13 estimated effect of the effect you believed, but the data
14 are the data. So, post hoc power analysis suggests that the
15 power is low.

16 Also just a comment, I appreciate and I also often
17 times involve myself in the imaginative work of
18 statisticians riding to the rescue of investigators who are
19 dealing with trials with compliance issues. But these are
20 investigator problems; they are not statistical problems. I
21 mean, investigators have to keep their patients on their
22 medications. That is why they randomized them. They have
23 to follow them to the end, and they have to ascertain vital
24 status and event status.

1 Any other solution is inferior. I mean, we, as
2 statisticians, can debate models on and on and no one or the
3 other will have any more basis in reality. It should be a
4 problem we shouldn't have to deal with, and the only way to
5 win this is not to play it. You know, keep the patients on
6 their meds. Investigators need to get that message clearly,
7 and follow patients to the end of the trial and be sure you
8 ascertain the appropriate event status of all these
9 patients.

10 DR. PACKER: I actually don't know whose problem
11 it is but sometimes it is something this Committee needs to
12 deal with actively because there aren't too many trials that
13 we see in which the compliance rate is 100%. So, the
14 question is there are fairly straightforward and I think
15 recognizably conservative approaches to dealing with the
16 issue of noncompliance if you have beaten the comparator,
17 for example placebo.

18 DR. MOYE: But we have to be careful too not to
19 let it slip away. Sometimes the investigators can get a
20 mind set that because there has been a statistical
21 adjustment for noncompliance it is okay if a few of my
22 patients go off medication. If that becomes infective, then
23 the trial really is no trial at all.

24 DR. PACKER: I understand, but let me follow

1 through on that because the theme you just brought up is a
2 theme that Rob brought up earlier, which is that there can
3 be many subtle influences on investigators, either non-
4 enforcement of compliance or many, many other aspects of the
5 trial which would blur and minimize true distinctions
6 between two treatments. Some of them, as Bob Temple said,
7 are so subtle that no committee, no matter how inquisitive
8 they may be, may be able to detect them.

9 I guess I am more concerned about when, in fact,
10 we can recognize that there is a problem, how do we deal
11 with it? In other words, we can't deal with distinctions we
12 can't detect, but when someone clearly presents to us a data
13 base in which there has been a noncompliance issue, how do
14 we deal with that, or do we simply say there is a problem
15 here and we have to mentally adapt our expectations
16 accordingly?

17 I guess, Dave, my question is I have already
18 gathered that there is no quantitative solution to the
19 problem of noncompliance in an equivalence-directed trial.
20 That is a correct statement?

21 DR. DEMETS: I don't know how you are asking it.
22 I can add to the quantitation of the problem. I mean, you
23 could, for example, suppose that the trial didn't figure out
24 how much noncompliance they would have, you have a result

1 that is non-significant, didn't meet the criteria, you could
2 go back and say, well, given this number of the
3 noncompliance rate what is the probability that I could have
4 found a difference in them anyway? It is after the fact but
5 you can get some sense of how bad off you were and what you
6 could have expected. It is not going to rescue the problem.

7 DR. PACKER: So, what will eventually be dealt
8 with since there is no quantitative solution is simply a
9 lack of individual conviction that the conclusions are valid
10 as stated. Is that fair?

11 DR. DEMETS: I think so.

12 DR. CALIFF: I just want to balance or maybe
13 disagree on this issue of questioning the investigators too
14 hard on compliance because one of the problems that we
15 frequently see in this regulatory process is selection of
16 ideal patients. I mean, we know that when we deal with real
17 patients in the real world there are all kinds of things
18 that happen to people, and reasons why they stop taking
19 their medicines, and that represents what the treatment is
20 really going to do when you go to prescribe it to the next
21 patient. So, we end up with these studies of professional
22 clinical trial patients that exist now, who will take their
23 medicines and give you beautiful dose-response curves but it
24 is not telling you about the effectiveness of that treatment

1 when it is going to be let our in the world.

2 So, although I agree that we have to do everything
3 possible to try to keep people on therapy, I would hate to
4 see that over-interpreted to mean that we want to pick
5 populations that don't overlap at all with the people that
6 we are actually going to have to treat when the product is
7 on the market.

8 DR. MOYE: As long as you and I agree that
9 investigators shouldn't use the statisticians as a crutch
10 for excusing patients from compliance requirements or vital
11 status ascertainment, you and I are in agreement.

12 DR. PACKER: I think there is general agreement on
13 that.

14 DR. LIPICKY: Rob, why do you blame that on the
15 regulatory process?

16 (Laughter)

17 DR. CALIFF: It is clearly an interpretation of
18 the regulatory process that you agree with but which is not
19 accepted by the people that are dealing with it.

20 DR. LIPICKY: No, no, no. No one agrees with
21 that, that I know about. But it is not dictated it has to
22 be otherwise.

23 DR. CALIFF: Then why is it that we have so many
24 studies that --

1 DR. LIPICKY: I have no idea.

2 DR. CALIFF: Well, it would be great if this
3 Committee could have some more direct communication with
4 people to try to get studies that represent the real people
5 that we are going to have to treat.

6 DR. LIPICKY: But I imagine it is in part related
7 to that variance term and the sample sizes that would be
8 necessary to show a difference of X, and the fear that
9 people have that the variance term would go up.

10 DR. CALIFF: Right, so --

11 DR. LIPICKY: But no one knows that it would, nor
12 has anyone, to my knowledge, demonstrated that that is true.

13 DR. CALIFF: So we end up with beautiful
14 experiments in populations which don't represent the people
15 we are going to have to treat so that we can reduce the
16 variance, at least in theory. That seems to be what is
17 happening.

18 DR. LIPICKY: But you agree to do trials of that
19 sort. So, don't blame it on the regulators.

20 DR. PACKER: I think what Ray is saying is that
21 the choice of the type of study is in the hands of the
22 investigators and the sponsor.

23 DR. LIPICKY: Yes.

24 DR. PACKER: And he is generally receptive to any

1 data submitted to him.

2 (Laughter)

3 DR. PACKER: Is that true, Ray? I didn't say that
4 you would like the data but you do receive the data.

5 DR. LIPICKY: That is correct, yes.

6 DR. CALIFF: But, Milton, there is a difference
7 between receiving and encouraging worthwhile studies. Those
8 are two different things.

9 DR. PACKER: We will get into that in one second.
10 Hold on. Ralph?

11 DR. D'AGOSTINO: The notion of not encouraging
12 noncompliance and so forth, obviously you have to agree with
13 that but there are statistical ways of making adjustments
14 that are beyond just superiority trials. I mean, you know,
15 you can look at the superiority trials and the adjustments
16 you make are on the conservative side. You allow an
17 adjustment that is going to make it hard to show the
18 superiority, and when you move endpoints forward and so
19 forth, you do it if it is going to work against showing what
20 you want. You can play the same game with the equivalency
21 trials and allow adjustments that are going to make it hard
22 to show the equivalency. You know, a lot of statisticians
23 will have made their careers on imputation and so forth, and
24 those techniques will come more and more in these

1 equivalence trials.

2 I think the interpretation though at the end of
3 the day is extremely hard, and to encourage that after the
4 fact you can make these adjustments is really not the
5 appropriate way to do it, but there are ways of doing it.
6 To answer the original questions, there are statistical ways
7 of making those adjustments and you can see just how bad the
8 compliance and noncompliance actually impacted on your
9 results.

10 **General Discussion**

11 DR. PACKER: Maybe we should let Dave sit down
12 before we open this up for general discussion, unless the
13 Committee has any other questions specifically for Dr.
14 DeMets.

15 We are supposed to have a general discussion but
16 we have already been having a general discussion for quite
17 some time. I thought that what might be useful as a
18 conceptual model for discussion for the remaining time
19 allotted to this session is -- Ray, let me postulate a
20 hypothetical, but presumably common scenario.

21 Before doing that, let me ask those who are in the
22 audience, how many of you are thinking about or are doing an
23 equivalence trial?

24 (Show of many hands)

1 Let me ask, how many of you were thinking about
2 doing that before today?

3 (Laughter and show of few hands)

4 Okay. Ray, when a sponsor comes to you and says I
5 want to do an equivalence trial, what do you say to them?

6 DR. LIPICKY: Go away.

7 (Laughter)

8 DR. PACKER: Then it is a pretty short meeting?

9 DR. LIPICKY: Yes. Do you want a longer answer?

10 DR. PACKER: Well, I just want to know if there
11 was further discussion and what it generally consisted of.

12 DR. LIPICKY: Well, the general discussion sort of
13 is on the lines of the general discussion today. It is what
14 area are you thinking about? What positive control are you
15 thinking about? How will you, for that positive control in
16 this area, be able to get an estimate of the effect size of
17 treatment? Because without some estimate of that effect
18 size it becomes rather difficult to talk about how much of
19 the effect must be preserved. Then you need to think about
20 how much of the effect needs to be preserved, and develop an
21 argument for that. Then you have general ball park
22 estimates for how you begin to calculate sample size because
23 you then know what the variance is and all that sort of
24 stuff.

1 DR. PACKER: Do you tell them they need one or two
2 equivalence trials?

3 DR. LIPICKY: Two.

4 DR. PACKER: Does it matter how persuasive -- I
5 understand you tell them but what do they say?

6 DR. LIPICKY: They say we will only do one.

7 (Laughter)

8 DR. PACKER: Okay, I understand. Bob?

9 DR. FENICHEL: Actually, we do see equivalence
10 trials, not as the proposed basis of approval but we
11 certainly see trials that could be interpreted as
12 equivalence trials all the time when someone with, let's
13 say, an antihypertensive will do placebo-controlled trials
14 showing it lowers blood pressure, and so on, and then they
15 will do some sort of trial where they run against some
16 popular antihypertensive and show, well, the effects are
17 kind of the same.

18 It is accepted that the results of that kind get a
19 somewhat vague but -- you know, they get some words into a
20 statement into the labeling that say in trials where they
21 used this and they also used nifedapine or they also used
22 hydrochlorothiazide, or whatever, the results were kind of
23 the same. It is very vague. It is not a real claim. It is
24 something which we keep people from promoting as a claim,

1 using in advertising or anything like that, but people seem
2 to like it so we let them do it. If someone wanted a strong
3 comparative claim, saying that this is better than
4 nifedapine, then we have a fairly rigid rule of two trials.

5 DR. LIPICKY: Well, it is like the usual
6 conversations. We were both talking about two different
7 things. If one is talking about, say, some morbid mortal
8 trial where up front the event rates are relatively low, and
9 where the original claim is where you give drug X and you
10 then change irreversible events, that is a little different
11 from the business where you are just sort of playing around,
12 and you can play around as much as you want and don't get
13 into any trouble if you play around as much as you want even
14 though you get no information. So those are two very
15 different things.

16 Indeed, for an antihypertensive to attempt to make
17 an equivalence claim or a superiority claim, we just went
18 through that exercise yesterday or the day before yesterday,
19 something like that, and there were five people in the room
20 and all five people had different ideas. We eventually told
21 the company something, but nobody said the same thing. The
22 problem there, obviously, is the problem that has been
23 discussed. If you compare 1 mg of nifedapine to 100 mg of
24 enalapril once a day, those are going to have different

1 effects, but it has nothing to do with whether or not the
2 drugs actually have a different effect.

3 So, it is not a chemical claim. Then the argument
4 sort of comes down to, well, we don't regulate chemicals; we
5 regulate dosing regimens. So, then it is possible to say,
6 well, one can compare one dosing regimen to another dosing
7 regimen and say this dosing regimen is better than this
8 dosing regimen or is equivalent to the dosing regimen. It
9 becomes a very hairy, complicated problem that is even more
10 difficult than the one we are talking about today.

11 DR. PACKER: Let me just pursue the hypothetical
12 scenario -- Dr. Seigel?

13 DR. SEIGEL: I also want to address that scenario
14 a little bit because we are also facing that, particularly
15 with a number of companies coming in with new thrombolytic
16 agents where they are generally not doing placebo. I would
17 say that in general the conversations follow the same gist
18 as Ray's conversations do.

19 However, once we have reached the point where we
20 are talking about trials, as we are typically of, say,
21 25,000 people, we then have yet to broach the issue of how
22 many of those are required.

23 DR. PACKER: I would imagine, as in the case of
24 superiority trials, it would depend on how persuasive one

1 trial was.

2 DR. SEIGEL: Right.

3 DR. PACKER: I guess one very good superiority
4 trial can be persuasive as one very good, appropriately
5 sized and outcome-dependent equivalence trial could also be
6 persuasive. But it would sound like it would have to be
7 very large to be persuasive.

8 DR. SEIGEL: Well, we have been using, and we will
9 be bringing this by the Committee at future points in time,
10 relative conservative answers to a lot of these questions,
11 how to estimate the effect size, and we look at the meta-
12 analysis but we look at the lower confidence intervals of
13 the meta-analysis. Those were done in trials where the
14 absolute effects were large, say, 2% and 8% mortality. Now
15 mortality is lower. Either they are lower risk populations
16 or the impact of aspirin and revascularization procedures
17 may lower the impact of thrombolytics. We don't know what
18 the effect size of thrombolytics are.

19 We use a relative as opposed to an absolute
20 difference as a more conservative approach. We require that
21 some of the effect be changed. We get into a lot of debates
22 about what the right control should be, and if the standards
23 are different depending on which control you choose. But
24 there are a lot of complexities to the design, and depending

1 on what assumptions and what approaches to the many issues
2 that were discussed here, you come out with very, very
3 different approaches and, therefore, I think the conclusion
4 you drew is right, that is, a single trial can be a very
5 powerful statement. Then you have the whole issue of is
6 there going to be good compliance, and the population. Is
7 it going to be done in people where the effect is large,
8 within the first 6 hours with S-T elevation? Or is it going
9 to be done in a setting where there is not much drug effect?
10 So it is going to depend on how you do that trial. But we
11 assume that in a very large multi-center trial it is likely
12 to be at least representative. The number of trials is less
13 important than the weight of the evidence.

14 DR. LIPICKY: Milton, I apologize. I was not very
15 responsive to the question you asked me and the way the
16 other people were responding reminded me of that. Indeed,
17 it isn't a one-trial, two-trial question. It is a question
18 of how persuasive the single trial is if it is a single
19 trial or if it is two trials. I think any single trial
20 could be as persuasive as you wanted it to be and that, in
21 fact, is the advice we give to people, that if they are
22 thinking that they are going to only be able to pull off one
23 trial, when they do their power calculations they ought to
24 figure that they are not shooting for 95% confidence limits

1 but they are shooting for some other confidence limit; and
2 that they should not in the slightest under-power their
3 study.

4 The general framework of reference we give is that
5 nobody would feel uncomfortable having two repetitions at a
6 0.5 level that sit in the same part of the tail of the
7 distributions, and that is the equivalent p value of
8 0.00125. You notice a change from 0.0025? When you see a
9 result that can convincing, you know, you are fairly
10 comfortable that that is real.

11 Now, the degree to which one is from 0.00125, and
12 this is not the p value but this is just being used for the
13 sake of talking, the degree to which you are closer to a p
14 of 0.5 than 0.00125 is the degree to which you have a less
15 powerful argument when you are looking at one study. People
16 usually talk about one study, two studies and p values.
17 Indeed, the proper way to look at it is in the light of the
18 difference between 0.05 and 0.00125.

19 DR. PACKER: When someone comes and wants to do
20 such a trial, do you tell them the goal is to demonstrate
21 that they are better than a putative placebo, or is the goal
22 to provide a reliable estimate of a treatment effect against
23 an active comparator?

24 DR. LIPICKY: Well, they accomplish both ends with

1 an appropriately designed positive control trial. Rob gets
2 all the information he wants and we can say it is better
3 than placebo so we get the information we want.

4 DR. PACKER: Rob, are you happy with that?

5 DR. CALIFF: No.

6 DR. LIPICKY: Why not, Rob?

7 DR. PACKER: Do you have a longer answer than
8 that?

9 DR. CALIFF: I am trying to imitate Ray.

10 (Laughter)

11 Because I don't think that there is enough active
12 encouragement going on right now to get people to do large
13 trials. By that, I mean if you flip the question around and
14 say let's assume we all agree that we would like to have
15 reliable estimates of what a new treatment does -- that is
16 what we all really want; that is what the public wants and
17 that is what patients want. The question is what are we
18 doing as leaders in this regulatory agency to take away the
19 impediments that exist to doing the size trials that we
20 need? I think there is a passive acceptance of sort of if
21 you do it, that would be great. But I don't see an active
22 effort being made to take away the impediments.

23 In fact, if you look at international trials,
24 which are generally required now to generate the kind of

1 sample size we are talking about, I see things actually
2 headed in the wrong direction in terms of diversion of
3 resources away from larger sample size and into regulatory
4 requirements that call for detailed reporting that costs a
5 huge amount of money and, as I said, auditing of data
6 because people don't believe that doctors tell the truth.

7 DR. TEMPLE: We are certainly seeing more drug
8 companies sponsor large trials, that is 10,000 or more, than
9 we ever have in the past. So, if there is much
10 discouragement, there is some other incentive out there that
11 overcomes that. But we should probably talk about whether
12 we are not helping as much.

13 I wanted to go back to what Milton said. I think
14 the dichotomy you have placed, that is, do you want to know
15 you beat placebo or do you want to have a good comparison of
16 the drug is fundamentally a false dichotomy. You can't even
17 begin an active control trial that doesn't show superiority
18 until you are quite certain the active control can beat
19 placebo reliably. So, if you can't be sure of that then
20 failing to see a difference, no matter how exquisite the
21 confidence intervals, is completely uninformative. You just
22 don't know whether the trial has any capacity to show
23 anything. So in a comparative setting you are as bound to
24 the need to have an active control that has a definable

1 difference from placebo as you are if your main difference
2 is trying to show a difference from placebo. You cannot
3 escape that. If the historical assumption that the control
4 will beat placebo isn't valid, you can't learn anything.

5 The question that then follows is how close do you
6 want to be, which you can define any way you like -- how
7 much of the placebo effect you want to preserve; or how much
8 of a difference between the two therapies in a setting where
9 the study is informative do you want to maintain? They are
10 not separate. They are together, and the second is a
11 judgment, how much of the effect do you want to preserve.

12 The first question, is this a trial where you are
13 quite sure that a placebo, had it been there, could have
14 been distinguished, you cannot even initiate a trial until
15 you know that. It is a nonsense, foolish trial because it
16 won't be informative. So, I don't think there is really a
17 distinction between those two things.

18 DR. CALIFF: Having made my statement before, I do
19 want to come back and say I agree completely with what Ray
20 said and also what Bob said. I mean, to do the minimally
21 important difference trial in this environment requires that
22 you have reasonable evidence that you are going to be better
23 than placebo. By the nature of minimally important
24 difference determinations compared to active controls, that

1 is an assumption. So, going through that exercise and
2 calculation is critical and I agree completely with it.

3 I just continue to push for more active
4 encouragement of the sample sizes that we are saying we
5 need, taking away impediments.

6 DR. COLLINS: I just want to comment that I think
7 it is important not to try to turn the argument around and
8 say what can we get away with in order to get approval. We
9 need to come back to this point about the aim being to get
10 reliable evidence that the new treatment is effective. Just
11 because we conclude that it is very difficult to do,
12 shouldn't then say, okay, let's make things look a bit lax;
13 let's make things a little bit easier; let's allow big
14 differences to be interpreted as equivalent. I think we
15 have to recognize that true equivalence studies are very
16 difficult to do; that the statistical uncertainty will mean
17 that they need to be much bigger; the clinical uncertainty
18 will mean that they are very difficult to interpret. But
19 you are not to say, well, what can we get away with? I
20 think we just have to recognize that that is the case. To
21 reinforce the point that if that is the circumstance, if you
22 cannot avoid doing a positive control equivalence study --
23 and I do believe that there are a lot of situations where
24 add-on studies could be done where they are not being done

1 and where they would be much better for society and would
2 actually be much better for the sponsors because they would
3 be easier to do -- but if we really can't avoid doing those
4 studies, then we need to actually make it easier to do them
5 and the point that Rob makes is absolutely correct. The GCP
6 guidelines are a major obstacle to achieving those ends
7 because they are all about accurate data points and not
8 about reliable answers. The philosophy underlying those
9 guidelines is completely wrong.

10 DR. SEIGEL: I have worked hard on that very
11 issue. The International Harmonization Process which, as of
12 three or four years ago, had a draft guideline which was
13 going to perpetuate the problem. And I am pleased to say,
14 although perhaps not yet fully reflected in federal
15 regulations, that the ICH International Conference on
16 Harmonization final guideline on good clinical practices is
17 very clear and explicit about the fact that the necessary
18 amount of monitoring is clearly a function of the intent of
19 the trial, the size of the trial, the design of the trial.
20 In several places it specifically accommodates the notion
21 that larger trials which collect more data on critical
22 endpoints, with less monitoring or with sampled monitoring,
23 may well be desirable and should not be excluded in any way.
24 So, there is an art of compromise here and the language

1 isn't as explicit in some places as we might like, but that
2 guidance, and one that I assume our regs are to come into
3 compliance with, is rather flexible. Our current regulatory
4 approach has been rather flexible on that issue as well.

5 I would like to say regarding a related issue,
6 which was raised by Rob Califf regarding poor compliance
7 perhaps reflecting reality and not, therefore, being a valid
8 setting in which to collect data, some amount of data
9 regarding safety is probably best obtained in an area where
10 great attention is paid to the level of compliance since the
11 effect size is rather important. A physician and a patient
12 don't need to know that a certain side effect is rare
13 providing you are like everybody else and don't take the
14 drug. They really need to know the effect size for safety
15 and to some extent for efficacy on the presumption that they
16 might take the drug. So, there is a balance between
17 information, I think, as to the true drug effect and
18 information as to what its effect will be in a true
19 situation where compliance may be poor.

20 DR. COLLINS: I am sorry, I have to differ on this
21 because I am not aware of any people who are actually doing
22 large-scale trials who were involved in developing the GCP
23 guidelines. And if you actually read them, they do not put
24 adequate emphasis on getting reliable answers. There is a

1 lot in them about getting accurate data points, and t here
2 is very little in them about getting reliable estimates of
3 the effects of treatment. And vague terms tend to be
4 interpreted to the maximum. So, sample monitoring means 90%
5 instead of 100% rather than sort of 1%. So, making it vague
6 actually doesn't help because it tends to be over-
7 interpreted by the supporters of a large number of studies,
8 which is industry. I mean, they want to make sure that they
9 don't get damage when they then go with the result.

10 DR. SEIGEL: Well, obviously it isn't black and
11 white, and I am sure isn't what you would like it to be nor
12 even perfectly what I would like it to be. Suffice it to
13 say that certain earlier versions of that document, as well
14 as earlier regulations in certain parts of the world
15 included language such as every data points needs to be
16 monitored, and every site needs to be monitored before,
17 during and after the trial. That sort of language is not
18 there. Instead, there is language that is vague but allows
19 for flexibility that monitoring used to be appropriate to
20 assure the quality of the data, and that is ultimately the
21 sponsor's responsibility, and it may well be that that will
22 be interpreted overly cautiously. I think that can only
23 come from ongoing dialogue, probably not from broad
24 guidelines. The Agency has been in dialogue with industry

1 and co-sponsored with PHARMA, in fact, a year and a half ago
2 a conference on data quality assurance. Rob and I, I think,
3 co-chaired a session on that. We were discussing
4 specifically large sample trials and the implications.

5 DR. FENICHEL: I just wanted to add something to
6 Ray's answer about what we say when people propose trials in
7 this area. That is, one of the things that I think is
8 important that we say is that you probably will, in many of
9 these areas, get one chance because if you do a small trial,
10 an under-powered trial, it may be sufficient to make it
11 impossible to recruit for any subsequent trial of this agent
12 because everyone may be convinced, on the basis of a finding
13 that is not even 0.05 but is 0.2, well gee, it sounds good
14 and my patients are really sick, and so forth and so on.
15 So, the game is over. And the idea of following on once you
16 have a kind of good idea this is a good place to put your
17 money, that may not be a realistic expectation. So, it is
18 appropriate to bite the bullet and say we are going to go
19 for a genuine finding of hard data, which means a trial of
20 adequate power, as we have heard described by Dr. Collins.

21 DR. TEMPLE: It is possible that in addition to
22 the GCP document, which is actually now in our regulations I
23 believe, we need to have some explanations that clarify some
24 of those things.

1 Jay and I both make real pests of ourselves on the
2 subject you are talking about. The current guidance, for
3 example, says that in some cases there may be no need for
4 on-site monitoring at all, which I can assure you is not
5 remotely what the document said initially. Although it is
6 true that there were no large simple trialists, so to speak,
7 in the room, we were very strong on explaining to the people
8 doing those guidances that almost all the really useful
9 information that had been generated over the years, at least
10 related to survival, came from trials that weren't monitored
11 like they were asking for.

12 So, the document leaves considerable room, and if
13 it is being misinterpreted we could probably develop some
14 clarifications. That might be useful if that is what you
15 are finding.

16 DR. PACKER: It would be fair to say, Bob, that
17 the adjective "simple" referred to the trial, not the
18 trialists?

19 DR. TEMPLE: Oh, yes.

20 DR. SEIGEL: Also "large."

21 DR. TEMPLE: Also "large." Most of them are of
22 average size, I would say. Can I say one other thing? One
23 of the issues that hasn't come up is stopping trials early,
24 which is a way to assure that you don't get extreme levels

1 of statistical significance.

2 One of the things we have been telling people, and
3 I would be interested in hearing comment on, is that while
4 there is some urgency to monitor a trial and stop it if you
5 are seeing a survival effect, there is less urgency when you
6 are doing a trial with a combine endpoint. So, we have been
7 encouraging people to stop only for survival outcome.

8 Another possibility, which I don't believe anybody
9 does but which is sort of consistent with what I understand
10 to be British practice, is to tell people at the outset that
11 you are only going to stop when you have a fairly extreme
12 view so that it is part of informed consent, and then keep
13 going until a very robust value is reached, which also
14 allows for the potential of exploring a subset of hypotheses
15 and things like that. But stopping early is a real menace,
16 especially when you are only going to get to do the trial
17 once.

18 DR. PACKER: Bob, I think that in practice the
19 only thing that can be reasonably monitored by an ethical
20 committee in an updated fashion would be mortality in
21 addition to the reasons that you have specified, and in
22 addition to the clinical persuasiveness because in most
23 cases non-fatal endpoints need to be adjudicated and that
24 induces delay.

1 DR. TEMPLE: Well, Milton, we are seeing people
2 who are getting good at this and who are doing that
3 adjudication sort of on the spot very rapidly. So, it will
4 actually be more possible than it has been to stop for a
5 variety of endpoints.

6 DR. CALIFF: I like your suggested approach of
7 stopping for mortality only, and we have had a recent
8 example where it was stopped for a composite endpoint, and
9 six-month data actually became clinically very important
10 and, of course, it wasn't powered to see a difference at six
11 months because the study stopped early for an extreme result
12 in the 30-day outcome. So, I think practical experience
13 perhaps leads to the same conclusion.

14 Again, you know, given the discussion, I just want
15 to reiterate that I know that Jay and Bob have both
16 struggled valiantly with some extreme bureaucracy. I think
17 there are large, simple trialists -- they favor large,
18 simple trials, I should say, given the discussion. But
19 vague documents done through committee, given the feelings
20 of this particular group, may not be adequate to do what is
21 in the public interest. It may be that a more explicit
22 statement, particularly in cardiovascular disease, would
23 encourage people to channel their money into larger sample
24 sizes and valid endpoints rather than more data points that

1 can be monitored by people flying around in airplanes.

2 DR. THADANI: Accepting that active trials will be
3 done against an active control and, as Rory said, you would
4 need a sample size of 50,000, 60,000 or maybe 100,000, what
5 reliance can we put on sample sizes which are 7,000, 10,000
6 or 20,000? Are we going to be able to look at them, or will
7 the FDA say, well, we are going to bring this trial for the
8 Committee to review because you can't make any judgment?
9 So, does the FDA tell the sponsor to do a trial of 80,000 or
10 100,000 otherwise we are not going to look at it? Or, if
11 you look at it, it has no meaning? Ray, will you comment on
12 that?

13 DR. LIPICKY: I don't think it is a sample size
14 problem. It is a power problem. You know, what is the
15 event rate? What is the difference that you are looking
16 for? What kind of acceptable difference would there be,
17 etc., etc., etc? It is not how many people you need. It
18 turns out that for most of the drugs that we are seeing
19 positive control trials for, the effect size is fairly
20 small. Consequently, you need a fairly large sample size to
21 talk about small effects. You know, if a trial clearly is
22 totally inadequate the Committee doesn't get to look at it,
23 I assure you.

24 DR. TEMPLE: As far as allowing people to carry

1 out trials, it would be unusual for us to stop a trial
2 because we don't think it was big enough, at least partly
3 because all the calculations of sample size are based on
4 some estimate of what the effect size is, and it could be
5 bigger. So sometimes you luck out and you win.

6 We would only stop a trial if we thought that it
7 was just inadequate by design to reach its goal, and that is
8 quite unusual. We are authorized to do that by our rules
9 but it is a very unusual thing to do. If a trial is
10 basically well designed we would say it doesn't seem likely
11 to get you anything; you are wasting your money, but we
12 wouldn't ordinarily stop it.

13 DR. PACKER: Let me just add one more question to
14 the hypothetical discussion with the sponsor. Bob Fenichel
15 earlier suggested that if A is better than B and B is better
16 than C, A may not be better than C --

17 DR. FENICHEL: No, I didn't say that.

18 DR. PACKER: Oh?

19 DR. FENICHEL: No, what I said -- and this is very
20 important so let me clarify this. What I said was that if A
21 is better than B and B is a lot better than C, one might
22 assume A is, therefore, a lot better than C. And that is
23 not true. The problem is that saying that something is a
24 lot better than something else is ambiguous. It may mean

1 that the effect size is very large, and it may mean that the
2 standard deviation is very small. So in one case you can
3 make this transitivity argument and in the other case you
4 can't. But, certainly, the much easier transitivity
5 argument, if A is better than B and B is better than C then,
6 sure, A is better than C.

7 DR. TEMPLE: In the same population.

8 DR. FENICHEL: In the same population. That is
9 true. What we have in the case of active controls, making
10 use of these combining arguments, is really not different in
11 kind but it is a little bit different from the arguments
12 that we make all the time. We have a body of, say, three or
13 four different trials all showing kind of the same thing,
14 that a drug lowers blood pressure or something, and now we
15 have to say, well, no one of these would be sufficient by
16 itself but we pool them together and say, yes, this is a
17 convincing argument that the drug works. Well, we are
18 assuming in that case that, gee, it is kind of the same
19 population, that these drugs are mutually reinforcing
20 because we are talking about some common biological
21 properties shared by the formulation given in each trial; by
22 the patients, they are all kind of the same species, and so
23 on.

24 That example of multiple trials in parallel, that

1 is a little easier because one of them could drop out and we
2 could decide, no, that was actually done in some other
3 species so we are not going to use that trial. Well, the
4 thing might still fly. When you have a bunch of drugs done
5 in series, which may be a fair description of A is better
6 than B, B is better than C and so on, then if any one of
7 them drops out the whole thing falls apart. So, there is
8 this assumption of a biological common substrate but, in
9 many ways, it is not different from what we face all the
10 time.

11 DR. PACKER: Bob, just a follow-up question. You
12 previously said that if a drug was better than an active
13 comparator, the choice of the active comparator might not
14 matter very much as long as you knew that the active
15 comparator was not harmful. Would that require a narrow
16 confidence interval? In other words, how does one know that
17 the drug that you have beaten is not a bad drug if it has
18 never been compared with placebo?

19 DR. TEMPLE: No, it would have to have been
20 compared with placebo. That is a data question. It also
21 goes partly to something that was said before. As Rory
22 pointed out or maybe Dave, the interpretation of a trial in
23 which you beat an active control is straightforward. It is
24 like interpreting a trial where you beat placebo once you

1 can make the assumption that the drug works.

2 So, a company that was trying to have a trial that
3 was easy to interpret would probably have an incentive to
4 use too low a dose or, you know, a drug that only works so-
5 so. That really wouldn't keep us from interpreting a trial
6 as showing effectiveness, but it might keep us if, say, the
7 dose was wrong from interpreting the trial as showing an
8 advantage over the drug. Those are two quite separate
9 things. You can do a trial to show that you work; you can
10 do a trial to show that you are better than something else.
11 And the two get kind of jumbled together sometimes.

12 DR. PACKER: Does anyone else on the Committee
13 have any comments or questions on any of the topics or to
14 any of our speakers today?

15 DR. RODEN: Bob Temple said something earlier
16 about IRBs and consent forms and differences between
17 American practice and U.K. practice. I just want to hear a
18 little bit more discussion about it. I mean, it is a
19 burdensome thing for an investigator to deal with an
20 industry-mandated consent form, which is what happens. If
21 one of the pleas that I have heard from Rob Califf is to
22 simplify things and to encourage large trials, it seems to
23 me that if we are going to make an investment in very large
24 trials and send all our money to Durham, North Carolina,

1 then at the very least we ought to make sure that the trials
2 that are conducted give us the best data possible. The
3 notion of including in a consent form the idea that the
4 trial won't be stopped unless some very, very hard endpoint
5 is reached has a certain appeal. Is there a mechanism that
6 we can use to encourage that practice?

7 DR. TEMPLE: There was an NIH conference several
8 years ago at which I remember throwing out the same
9 suggestion, but I have never heard any public discussion of
10 it. So, I don't know if anybody does that.

11 DR. RODEN: The Federal Register will get thrown
12 in your face if you try to change the IRB rules.

13 DR. TEMPLE: Well, this doesn't change an IRB
14 rule. This says that an IRB has to think up what is
15 appropriate. I mean, just as an example, if you have to
16 stop a study because of a one-month result and your real
17 interest is the three-year result, it is crazy. And it
18 isn't self-evident that you have to stop the trial if it
19 shows a significant difference. It depends. And I would
20 allege that an IRB can take those matters into account. It
21 does seem very important to make sure patients understand
22 what the drill is because they need to know the trial is
23 going to keep going on even though there was a survival
24 advantage and they might not like that. They might say they

1 don't want to be in a trial like that; they might say it is
2 okay with them. This is just a personal view. It hasn't
3 had widespread discussion, and it deserves it.

4 The reason I referred to a Transatlantic
5 difference is for reasons I am sure Rory can explain better
6 than I can. Many trials in the U.K. have gone much further
7 than would have been comfortable for some of the domestic
8 IRBs, and there has been some debate about that in various
9 journals. One of the points Rory made is that the point
10 here is to find out an answer that people are going to use.
11 It is stupid to get an answer that nobody is going to pay
12 attention to. You might as well not have bothered with the
13 study. That runs into some conflict with certain other
14 principles but it deserves a good discussion.

15 DR. PACKER: To summarize, there has to be
16 proactive and considerable amount of thought given to the
17 idea of what would constitute a persuasive result in a
18 trial, and that needs to be incorporated into the stopping
19 rules and, presumably, into the informed consent that would
20 allow a patient to be enrolled. That word "persuasive" is
21 not necessarily a nominal p value of 0.05. What is
22 persuasive will depend a great deal on the circumstances of
23 the trial, or its duration, or the endpoints that are
24 considered to be of clinical significance. I think that we

1 probably need to think carefully about what, in fact, would
2 constitute a persuasive result and adjust not only the
3 stopping rules of the DSMB but the IRB consent form
4 accordingly. The gold standard here is the persuasiveness
5 of a finding and not the mere existence of a finding. Would
6 that be fair?

7 DR. LIPICKY: This is maybe worth having a meeting
8 about because, you know, we make some recommendations, for
9 example, and we have had prestigious people, like Dave
10 DeMets and Tom Fleming, say that what we are asking people
11 to do is ridiculous but we are going to keep asking them to
12 do that. So, there is good reason to think that this kind
13 of stuff, which isn't talked about a lot, ought to have some
14 public discussion because it, in fact, does get in the way
15 and is a major problem in the conduct of trials.

16 DR. CALIFF: I read a very encouraging document on
17 the way up here, from the National Cancer Institute, which
18 indicates that the National Cancer Institute is going to try
19 to become a major force in simplifying this IRB consent
20 methodology in the United States, seeing it as a major
21 impediment to the public health. We are making it so hard
22 for people to participate in trials that we can't figure out
23 what treatments work. Maybe if groups like this got aboard
24 with the NCI and the regulatory agencies working with NCI we

1 could have a fairly persuasive group to foster change.

2 DR. PACKER: Ray, do you have any additional
3 questions?

4 DR. LIPICKY: No.

5 DR. PACKER: Any additional comments or questions
6 from our guests? Dave?

7 DR. DEMETS: I was going to respond to some of the
8 things that Ray attributes to me. If you say you are going
9 to do something in a trial and the protocol and the consent
10 form specifies it, then you get into this ethical dilemma of
11 having achieved what you said you were going to achieve but
12 you are not ready to stop. So my answer is to sort out what
13 you have just been implying. I think when asked about these
14 dilemmas, they go with the primary question in the protocol
15 and what the consent form says. That is the basis by which
16 they hold you, and if that is not what you wanted to do then
17 we should say so. That is my basic point.

18 DR. PACKER: Any other comments? Questions?

19 (No response)

20 I think we have had a very active discussion this
21 morning on all the issues. I think that in the analysis of
22 active control trials, clearly, if one is beating the
23 comparator the analysis is probably as straightforward as
24 the usual placebo-controlled trial. But if your goal is to

1 achieve a claim of equivalence the challenge is
2 substantially greater than I think many of us may have
3 previously imagined. That is an issue that I am sure will
4 be further explored as the number of trials which show
5 equivalence or claim equivalence are individually reviewed
6 in the future.

7 We will reconvene tomorrow at nine o'clock. The
8 Committee is having a closed session this afternoon. So, we
9 will convene in open session tomorrow at nine o'clock.

10 (Whereupon, at 12:00 noon, the proceedings were
11 recessed, to be resumed at 9:00 a.m., Friday,
12 October 24, 1997)